Internal redox labelling of oligonucleotides

Cabezas-Hayes, Sinead

Award date:
2018

Awarding institution:
University of Bath

Internal redox labelling of oligonucleotides

Sinéad Cabezas-Hayes

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Chemistry

June 2018

"All that is necessary for faith is the belief that by doing our best we shall come nearer to success and that success in our aims ... is worth attaining"

Rosalind Franklin

Table of Contents

Acknowledgements vii

Abstract viii

Abbreviations ix

1. Introduction 1

1.1. Pathogen Detection Methods 1

1.1.1. Immunoassays 1

1.1.2. Nucleic Acid Amplification Tests 3

1.2. Nucleic Acid Labelling Strategies 7

1.2.1. End-capping 11

1.2.2. Phosphate modification 12

1.2.3. Sugar modification 13

1.2.4. Non-nucleosidic units 16

1.2.5. Base labelling 17

1.2.6. Post-synthetic modification 20

1.3. Nucleic Acid Detection Methods 23

1.3.1. Radiolabelling 23

1.3.2. Colourimetric and fluorescent detection 23

1.3.3. Electrochemical detection 29

1.3.3.1. Surface-immobilised detection methods 30

1.3.3.2. Solution-based detection methods 35

1.4. Aims 38

2. Design, synthesis, and testing of mono-ferrocenylated nucleosides 40

2.1. Introduction 40

2.2. Nucleobase modification 46

2.2.1. Uridine substitution 46

2.2.2. Cytidine substitution 48

2.2.3. Purine bases 51

2.3. Ferrocenyl azide library development 53
2.3.1. Substituted ferrocenes 55
2.3.2. Conjugation of ferrocene to 2'-deoxyuridine *via* the copper-catalysed azide-alkyne cycloaddition 59
2.3.3. Phosphoramidite activation of redox active nucleosides 61
2.4. S1 digestions of mono-ferrocene oligonucleotides 64
2.4.1. Internal Control sequence selection 64
2.4.2. CT sequence selection 71
2.5. Conclusion 73
3. Design, synthesis, and testing of di-ferrocenylated nucleosides 74
3.1. Introduction 74
3.2. Di-ferrocene synthesis and nucleobase conjugation 77
3.3. S1 digestions of di-ferrocene oligonucleotides 88
3.3.1. Comparison of probe sensitivity 91
3.3.2. Mass spectrometry analysis 92
3.4. Conclusion 92
4. Internal Labelling in Diagnostic Assays 93
4.1. Introduction 93
4.1.1 Point-of-Care diagnostics 94
4.1.2 Atlas Genetics' Diagnostic Assay 95
4.2 Double stranded digestions of internally labelled probes 98
4.2.1 Internal Control double stranded assays 98
4.2.2 *Chlamydia trachomatis* double stranded assays 102
4.2.2.1 Mono-ferrocene *C. trachomatis* probes 103
4.2.2.2 Di-ferrocene *C. trachomatis* probes 106
4.2.3 Labelling patterns and mechanistic insight into T7 exonuclease digestion 111
4.3 Multiplex detections with internally labelled oligonucleotides 114
4.4 Lambda-exonuclease digestions 122
4.5 Conclusion 123
4.6 Summary 125

5. Experimental 126
5.1. General considerations 126
5.1.1. Equipment 126
5.1.2. Reagents 126
5.2. Biological Procedures 128
5.2.1. Primer sequences 128
5.2.2. Buffer solutions 128
5.2.3. Enzymatic assay protocols & mixtures 129
5.2.3.1. S1 Digestions 129
5.2.3.2. PCRs and synthetic target mixes 130
5.2.3.3. Double-stranded digestions 132
5.3. Chemical procedures and characterisation 134
References 191
Appendix 1 202
Appendix 2 203
Appendix 3 204
Appendix 4 205
Appendix 5 213
Appendix 6 236

Acknowledgements

I would like to thank my supervisor Prof Christopher Frost for giving me the opportunity to study and for his extended support over the last few years, where his continued enthusiasm in my project made it easy to enjoy my work from day one.

I would also like to thank Atlas Genetics for funding for my studentship and giving me the opportunity to present my work at an international conference. I would also like to thank several of their employees in particular; David Styles, Antony Brown, Dr Mike Storm, Ben Reynolds, and Dr Alistair Muir. Many thanks to you guys for teaching me how to biology, showing me where the *Taq* is for the 20th time, and for welcoming me into your working environment without hesitation.

Special thanks must go to Dr Barrie Marsh, for translating the biology speak and giving me guidance whenever I've asked. Nothing says welcome to postgraduate studies quite like handing me a large bottle of pyrophorics along with a fire extinguisher "just in case".

Thank you to the rest of the Frost Group I have had the pleasure of working with: Dr Sean Goggins, Dr William Mahy, Dr Andrew Paterson, Dr Jamie Leitch, Callum Heron, Sam Spring, Chi Zhang, and the many transient students we've shared the lab with. Being stuck in a tiny office with no windows isn't always easy, so thank you for making me laugh every single day.

My rugby teammates, of whom there are too many to name, have all played a part in keeping me sane(ish). From simply making me laugh after a bad day in the lab, mocking me relentlessly, taking me to hospital when I've broken something, to putting a roof over my head, I really don't know where I'd be without you.

Finally, thank you to my family, particularly my step dad, Andy, for putting up with me at home while I've been writing, and of course to my mum, Jenny. Thank you for your unfaltering advice, support, and belief.

Abstract

The aim of this thesis was to improve the detection method of electrochemical solution-based DNA diagnostics. Electrochemical solution-based detection lacks in sensitivity compared to surface-immobilised techniques or fluorescent probes, however benefits from cheaper equipment and materials, and lower levels of user-input required. Improvements in electrochemical signal upon detection of DNA would allow for improved sensitivity of the probes and therefore more useful diagnostic devices. To achieve this, it was hypothesised that the incorporation of several ferrocene labels onto oligodeoxynucleotides could offer improved electrochemical performance.

The synthesis of a library of ferrocene-based nucleotides is discussed in Chapter 2, with a view to increasing the possible increase in electrochemical signal analysed *via* differential pulse voltammetry that can be achieved *via* the introduction of multiple ferrocene units on a single strand. This was achieved through the conjugation of ferrocene onto 2'-deoxyuridine utilising the copper-catalysed azide-alkyne cycloaddition, followed by solid phase synthesis of oligonucleotides to allow the introduction of up to five ferrocenes on a single oligonucleotide strand. The digestion of two DNA sequences was studied with S1 nuclease, showing an increase in signal is possible in a single-stranded assay.

Chapter 3 expands on this methodology, next synthesising ferrocene labels with increased sensitivity *vs.* those developed in Chapter 2. This was achieved through the introduction of two ferrocenes onto a single nucleotide, followed by incorporation into a clinically relevant DNA sequence, allowing the synthesis of oligonucleotides containing ten ferrocene units. The probes were again treated with S1 nuclease displaying increased sensitivity measured *via* differential pulse voltammetry over the traditional 5'-labelling method, as well as the probes developed in the previous Chapter.

Chapter 4 details the incorporation of the probes previously developed into diagnostic assays, using both T7 exonuclease and lambda exonuclease to digest the novel probes. The internal labelling strategy was shown to be active towards T7 exonuclease digestion, however this did not offer improved digestion properties compared to traditional methods. The use of multiplex detection system was suitable for the detection of two clinically relevant DNA sequences simultaneously, while also offering insight into the mechanistic properties of T7 exonuclease. The incorporation of internal labels enabled the use of lambda exonuclease, the use of which has been limited by the previous labelling strategy.

Abbreviations

A	amperes
Ac	acetyl
AG	Atlas Genetics
AQ	anthraquinone
aq.	aqueous
Asc	ascorbate
B	generic nucleobase
Boc	*tert*-butyloxycarbonyl
bpy	2,2'-bipyridyl
BSA	bovine serum albumin
Bz	benzoyl
cDNA	complementary DNA
Cp	cyclopentadiene or cyclopentadienyl
Cp*	1,2,3,4,5-pentamethylcyclopentadiene or 1,2,3,4,5-pentamethylcyclopentadienyl
CPG	controlled pore glass
CT	*Chlamydia trachomatis*
CuAAC	copper-catalysed azide-alkyne cycloaddition
CV	cyclic voltammetry
dA	2'-deoxyadenosine
DABCYL	4-(4-(dimethylamino)phenylazo)benzoic acid
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
dC	2'-deoxycytidine
dG	2'-deoxyguanosine
DIPEA	*N,N*-diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DMF-DMA	dimethylformamide dimethyl acetal
DMTr	4,4'-dimethoxytrityl
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DPPA	diphenylphosphoryl azide
DPV	differential pulse voltammetry
dsDNA	double-stranded deoxyribonucleic acid
dT	2'-deoxythymidine
dU	2'-deoxyuridine
EB	ethidium bromide
ELISA	enzyme-linked immunosorbent assay
E_{ox}	oxidation potential
ESI	electrospray ionisation
equiv.	equivalents
FAM	6-carboxyfluorescein
Fc	ferrocene or ferrocenyl
FRET	Förster resonance energy transfer
h	hours
HIV	human immunodeficiency virus
HPLC	high performance liquid chromatography

HRMS	high resolution mass spectrometry
IC	Internal Control
i-Pr	*iso*-propyl
IR	infra-red
ITO	indium tin oxide
LOD	limit of detection
M	Molar
MB	methylene blue
MBG	molecular biology grade
m-CPBA	3-chloroperoxybenzoic acid
mins	minutes
MP	melting point
Ms	methanesulfonyl
MS	mass spectrometry
NAAT	nucleic acid amplification test
NBS	*N*-bromosuccinimide
n-Bu	butyl
NG	*Neisseria gonorrhoeae*
NMR	nuclear magnetic resonance
NTC	no template control
ODN	oligodeoxynucleotide
PCR	polymerase chain reaction
Ph	phenyl
PNA	peptide nucleic acid
POC	point-of-care
POCT	point-of-care test
p-TSA	*para*-toluenesulfonic acid
qPRC	quantitative polymerase chain reaction
RNA	ribonucleic acid
rt	room temperature
S1	S1 nuclease
sat.	saturated
SNP	single nucleotide polymorphism
ssDNA	single-stranded deoxyribonucleic acid
ST	synthetic target
STI	sexually transmitted infection
T7	T7 exonuclease
TAMRA	5-carboxytetramethylrhodamine
TBAB	tetrabutylammonium bromide
TBAF	tetrabutylammonium fluoride
TBAI	tetrabutylammonium iodide
TBS	*tert*-butyldimethylsilyl
t-Bu	*tert*-butyl
TEA	triethylamine
Tf	triflyl
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin layer chromatography
T_m	melting temperature
TMEDA	*N,N,N′,N′*-tetramethylethylenediamine

TMS	trimethylsilyl
TMSA	trimethylsilylacetylene
Ts	tosyl
U	1 U = amount of enzyme required to turnover 1 μmoles of substrate per minute
V	volts

1. Introduction

Sensing, in its broadest definition, is defined as the detection of an analyte to give a signal or response to its presence. The analyte varies from physical properties to chemical or biological substances, and the signal can be an optical, electrical, or physicochemical readout. Sensors are used in almost every aspect of modern life – computers rely entirely on them, as well as many medical devices being heavily reliant.

A biosensor is a device that uses biological components in combination with a label which gives a physicochemical readout. Biosensors have been commercialised to allow for the detection of chemical weapons and explosives,[1,2] waste water analysis,[3] glucose sensing in diabetic patients,[4] and bacterial and viral analysis[5], amongst many other uses.

Pathogen detection is important in many industries, including food production, clinical diagnosis, environmental analysis, and defence, amongst others. For public health, the detection focusses on food, water treatment (both supply and waste streams), and clinical settings for direct impact on the wider health of the population.

1.1. Pathogen Detection Methods

Conventional techniques for pathogen detection include microscopy, culture, and serology.[6] Culture remains the gold standard for the diagnosis of many pathogens. When conducted correctly it gives unequivocal results, however the growth of the desired pathogens are subject to potential failures and cross contamination. For some microorganisms, culture is not a suitable option due to either the difficulty or speed at which the organisms grow, or the associated danger with growing the organisms (*e.g*. *Treponema pallidum,* HIV, *Mycobacterium tuberculosis*, *Coxiella burnetii*).[7] Microscopy is operationally simple, however lacks sensitivity and results could be interpreted subjectively. The development of molecular recognition techniques allowed for the development of equally as reliable, but quicker and simpler to conduct methods of diagnosis, which will be detailed herein.

A biosensor relies on biomolecular recognition, and although the binding event of a target such as an antibody or protein is highly specific, they normally do not directly result in a measurable output without chemical modification. Biosensors relying directly on binding biological materials therefore usually require separation steps to ensure accurate readouts. Such techniques include Western blots and enzyme-linked immunosorbent assays (ELISA).

1.1.1. Immunoassays

ELISAs remain one of the most extensively used diagnosis methods. Its success draws from the specificity of antibodies, whereby the antigens are signalled by enzymatically labelled antibodies.[8] The most common form for the diagnosis of pathogens is a “sandwich” ELISA

(Figure 1.1), which captures the antigen between two antibodies, one of which is immobilised onto a surface. The second antibody is labelled with an enzyme, commonly horseradish peroxidase or alkaline phosphatase, which then reacts with the substrate to produce a signal. Any non-specifically bound materials are washed away, so unless the target is present, the enzyme will not turnover any signal. Extensive research into the field of ELISAs has led to the development of colourimetric,[9] fluorescent,[10] and electrochemical versions.[11] ELISAs have also been used outside the detection of pathogenic materials, most famously in lateral-flow assays for the detection of the hormone human chorionic gonadotropin for pregnancy tests. ELISAs can be used as standalone biosensors, or in combination with other molecular techniques such as PCR or signal amplification methods, leading to high selectivity and specificity for a range of analytes.

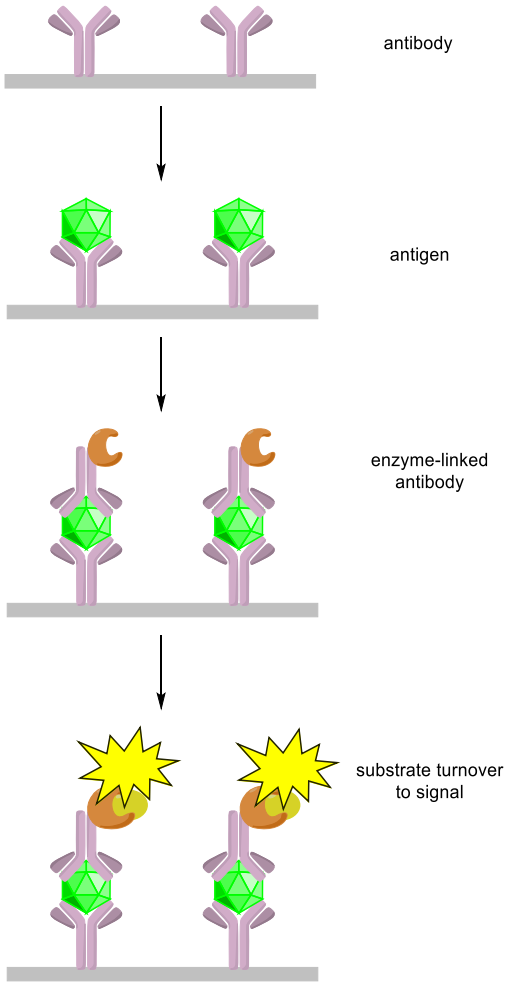
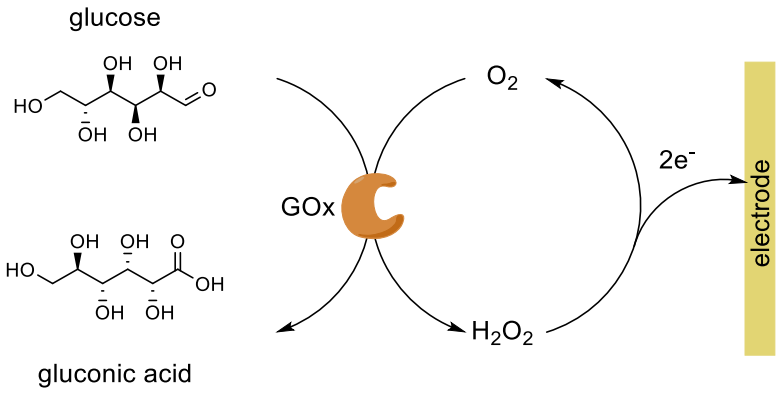


Figure 1.1 Capture "sandwich" ELISA

Western blots are a similar immunoassay, whereby proteins in a sample are separated by gel electrophoresis, then incubated with a labelled antibody. The label then reacts with a substrate to produce a signal (usually fluorescent or colourimetric) which can be detected by a device. Western blots are considered by many to be the gold standard for the detection

of HIV.[12] Although highly selective and specific, ELISAs and Western blots are rarely found outside of clinical laboratories, and as such, HIV and pregnancy tests are the only ELISAs to have been widely implemented in decentralised settings.

The only other widely implemented biosensor for medical applications is the home blood glucose monitor, used by millions of diabetics worldwide. The device uses the enzyme glucose oxidase, which catalyses the reaction of glucose into hydrogen peroxide, which is then oxidised at the electrode surface (Scheme 1.1). This produces an electric current which can be measured and is proportional to the serum glucose concentration, therefore allowing the user to quantify their blood glucose level. Home glucose monitors are still plagued by inconsistencies in results, with the International Organisation for Standardization (ISO) allowing for large margins of error compared to laboratory results.[13] ISO ensures commercial equipment are standardised across a given industry, updating its database to mirror chemical, biological, and technical advances.

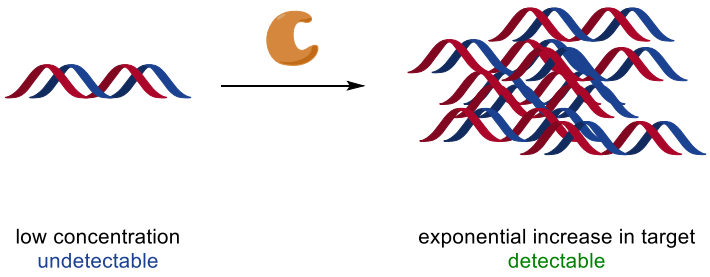


Scheme 1.1 Schematic of glucose detection in a home glucose monitor

These three techniques are all enzyme-based, exploiting the high specificity of the active sites. A major drawback of antigen-based immunoassays and non-amplification tests for pathogen detection is the window period associated with infection. For example, HIV can go undetected in a newly infected patient for up to three months, while during this window period the patient's immune system is developing a detectable level of antibodies. Until this point, the pathogen cannot be detected, therefore leading to many false negatives.

1.1.2. Nucleic Acid Amplification Tests

Nucleic acid amplification tests (NAATs) bypass the window period by amplifying low levels of pathogenic RNA or DNA to a detectable level, thus the pathogen can be diagnosed as soon as, or very soon after, the initial infection has occurred. NAATs are one of the most reliable and widespread techniques available for biomolecular diagnostics due to their reliability and unambiguity of the results.[14,15]



Scheme 1.2 Enzymatic amplification of nucleic acid

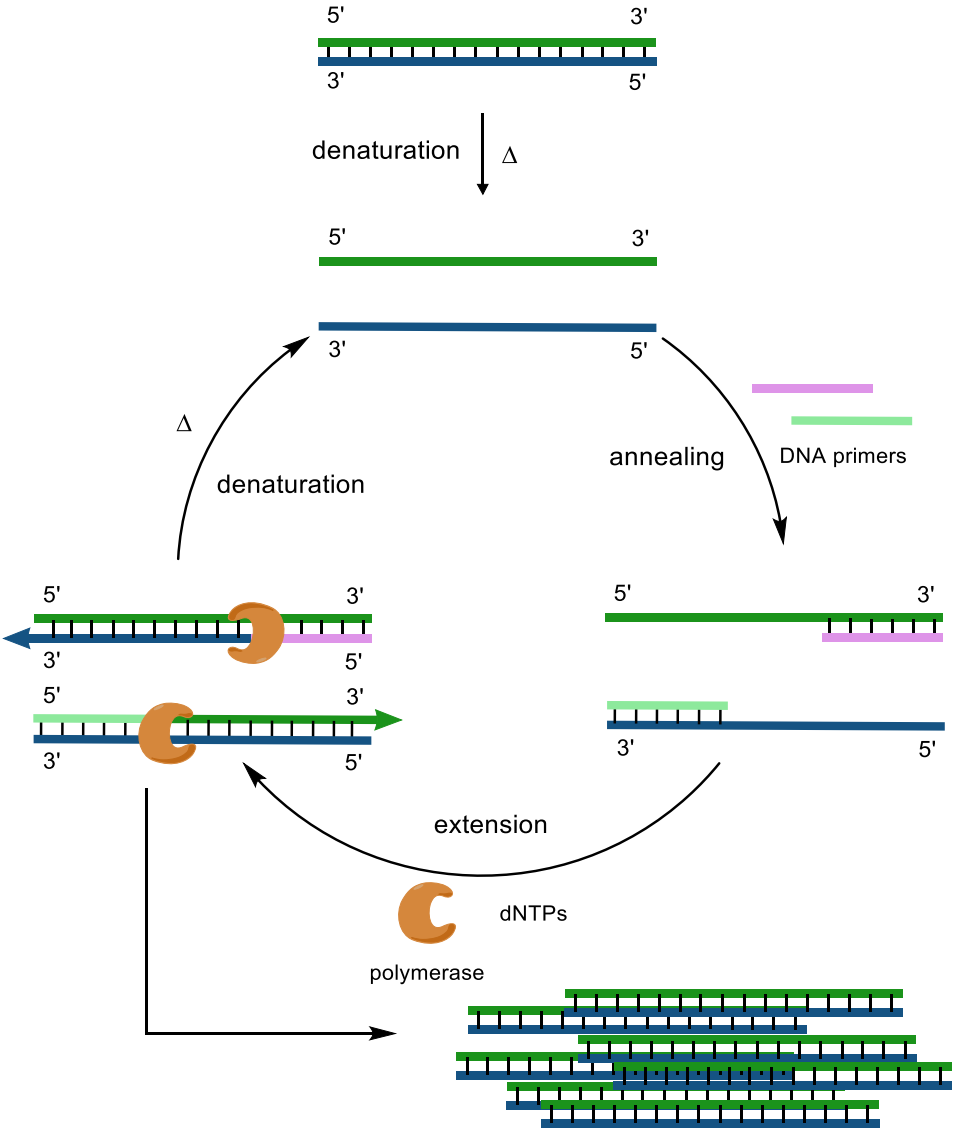
NAATs are commonly used for the detection of bacteria in both clinical and agricultural settings, with a range of research and commercial tests available for common foodborne infections such as *E. coli*, and sexually transmitted infections including *C. trachomatis* and *N. gonorrhoeae*.[16–19]

NAATs on clinical samples are conducted *via* three main steps: extraction, amplification, and detection. A sample is first collected and the target DNA must be isolated. For whole blood samples, this usually requires a purification technique, whereby the bacteria are separated from the rest of the complex matrix. For urine and swabs, this is a simple lysis, usually through salt/water treatment. By varying the osmotic pressure, the cells are forced to swell and burst, releasing genetic material. As bacterial DNA is not enclosed inside a nucleus, the DNA is released directly into the lysate. This DNA is then available to be amplified *via* amplification techniques.

The most well-known amplification technique is the polymerase chain reaction (PCR), which uses DNA polymerases to extend DNA primers. Other nucleic acid amplification techniques have been developed, but none of the isothermal techniques have reached the same level of success as the polymerase chain reaction.[20] PCR was developed in the 1980s by Mullis *et al*. and used to identify sickle cell anaemia.[21] PCR has since been extended for use in DNA sequencing, gene analysis, disease diagnosis, and pathogen detection.

In a standard PCR cycle (Scheme 1.3), double-stranded DNA (dsDNA) is first heated and denatured to give two strands of complementary single-stranded DNA (ssDNA). Upon cooling, short primers with complementary sequences anneal to the beginning of the target sequence, indicating the section of DNA to be amplified. Using the ssDNA as a template, a DNA polymerase then extends the two existing single strands *via* the incorporation of deoxynucleoside triphosphates (dNTPs), with the polymerase catalysing the nucleophilic attack of the existing 3'-hydroxyl of the DNA primer into the α-phosphate of the complementary dNTPs, yielding two new copies of dsDNA. The thermal cycling is repeated with the new dsDNA for the required number of cycles, such that each cycle effectively doubles the number of copies present (assuming 100% efficiency). The amplicons are

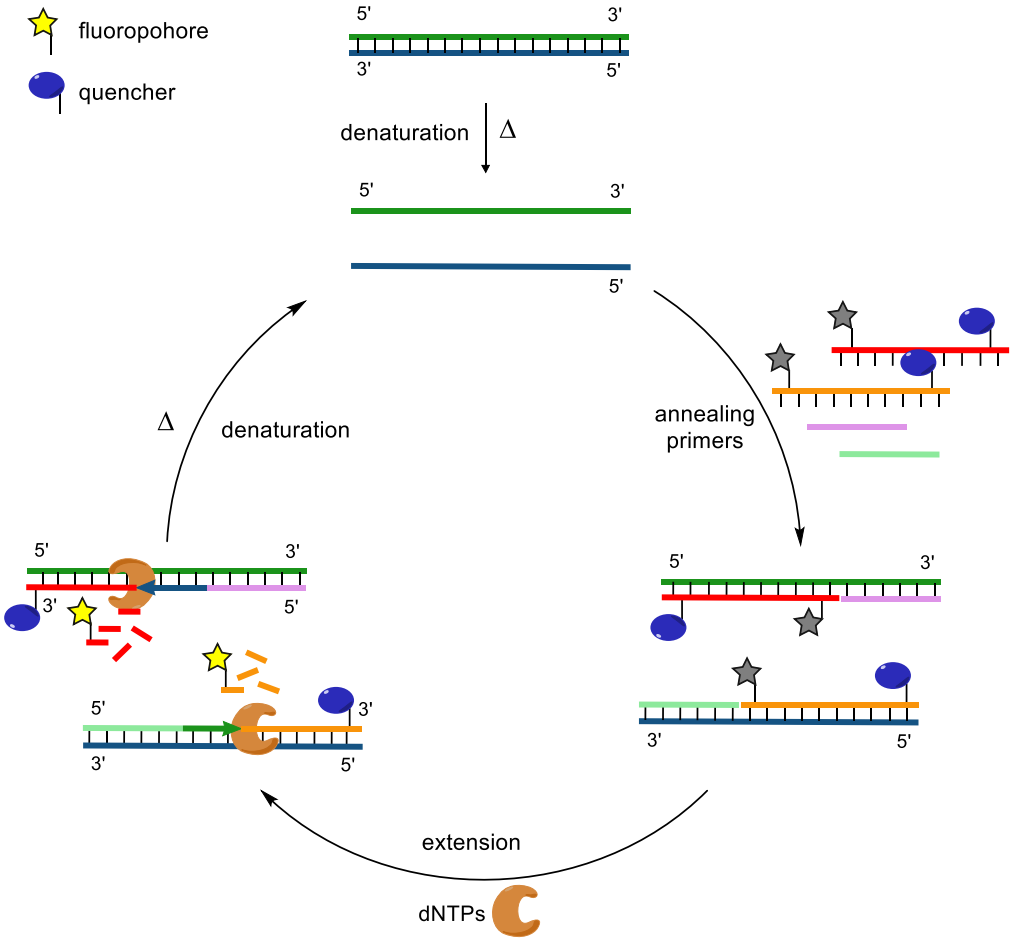
detected *via* gel electrophoresis, or are treated with labelled DNA probes to allow for specific detection of the amplicons.



Scheme 1.3 Polymerase chain reaction cycle

In the early days after the discovery of PCR, the experiments were time intensive and labourious. The enzymes first used were not stable upon heating which is an essential step to denature the DNA. This required the user to manually transfer fresh enzymes to the sample after each cycle, however the discovery of *Taq* polymerase, a thermostable DNA polymerase, led to the development of quicker and more reliable PCR methods.[22] When first invented, PCRs were performed using hot and cold water baths, which the user would have to move the PCR tubes between. Since then, driven by the discovery of *Taq*, thermal cycling blocks have become common practice, allowing the PCR cycling process to be fully automated.

Quantitation of DNA can be achieved through real time (quantitative) PCR (qPCR). This was first developed using ^{32}P labelled primers,[23] and later simplified with the use of fluorescent probes such as DNA intercalators or oligonucleotide probes.[24] In this variation, a second, fluorescent probe is added to the solution containing a fluorophore-quencher tethered to the same probe (Scheme 1.4).[25] Upon extension of the DNA primers, the fluorophore is released into solution due to the dual polymerase and exonuclease action of the enzyme used. The fluorescence is no longer quenched, thus resulting in a switch-on signal. qPCR has been used for applications where the amount of DNA is indicative of a genetic condition, or the success of treatment *via* gene expression (*e.g.* cancer diagnostics).[26,27] It has also been used to quantify bacteria such as *C. diff* and MRSA, and bacteria responsible for infections such as meningitis, pneumonia, and anthrax.[28]



Scheme 1.4 Real-time quantitative PCR cycle

Amplification tests allow for sensitive analysis of complex matrices. Coupled with modern detection methods, nucleic acid analysis allows for multiplex detection of biologically relevant samples with high precision and accuracy. Detection methods have advanced since the early days of DNA diagnostics, as have the synthetic methods used to functionalise oligonucleotides.

1.2. Nucleic Acid Labelling Strategies

Biomolecular recognition events (*e.g*. duplex formation) do not usually result in a measurable change in physicochemical properties. As such, a range of methods utilising different labelled probes have been developed to enable the detection of the recognition event. Detection and qualitative analysis of DNA has developed at a tremendous rate since the elucidation of the structure of DNA in 1953.[29] The structure of DNA is formed from three parts, the nucleobase, ribose, and phosphate backbone, together forming a single nucleotide unit. The phosphodiester linkages chemically bond single nucleotides to each other, allowing long fragments of DNA to be created. Watson-Crick base pairing (Figure 1.2) allows formation of a double helix, whereby adenine (A) pairs with thymine (T), and guanine (G) pairs with cytosine (C) *via* H-bond acceptors and donors located on the nucleobases, as well as van de Waals forces and π- π interactions.[30]

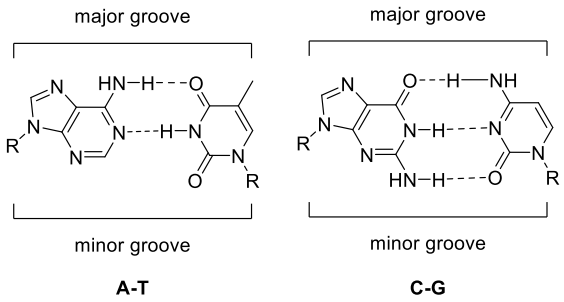


Figure 1.2 Watson-Crick base pairing of canonical nucleobases

The formation of the double helix of dsDNA creates two grooves, namely the major and minor, indicating the smaller and larger spaces between the phosphate backbone formed in the helical structure. The two strands of DNA are antiparallel, such that one runs 5'-3' while its reverse complement runs 3'-5' (Figure 1.3).

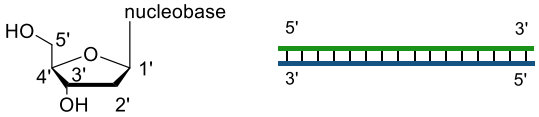


Figure 1.3 DNA numbering nomenclature

Nucleases are enzymes that catalyse the hydrolysis of the phosphodiester backbone to yield a hydroxylated and a phosphorylated section. Dependent on the specific nuclease, they can act in either a 5'-3'- direction or a 3'-5'- direction. Exonucleases cleave from the terminus of the oligonucleotide strand, whereas endonucleases cleave at internal phosphate groups. Most nucleases have a very specific set of conditions under which they will cleave DNA, with

selectivity for single or double stranded, DNA or RNA, restriction sites, and phosphorylation determining whether an enzyme will digest target DNA.[2]

The knowledge formed a basis of understanding of how DNA interacts with itself, and as such allowed for the intelligent design of DNA probes and sensors. Presently, biochemists have a range of labelling strategies available for detection methods, each with their benefits and drawbacks. These external labels are based on a range of functionality and can be chosen to suit the end-user's specific requirements. As such, labelling strategies of nucleic acids has drawn much attention, such that the selective introduction of labels can now be achieved with relative ease.

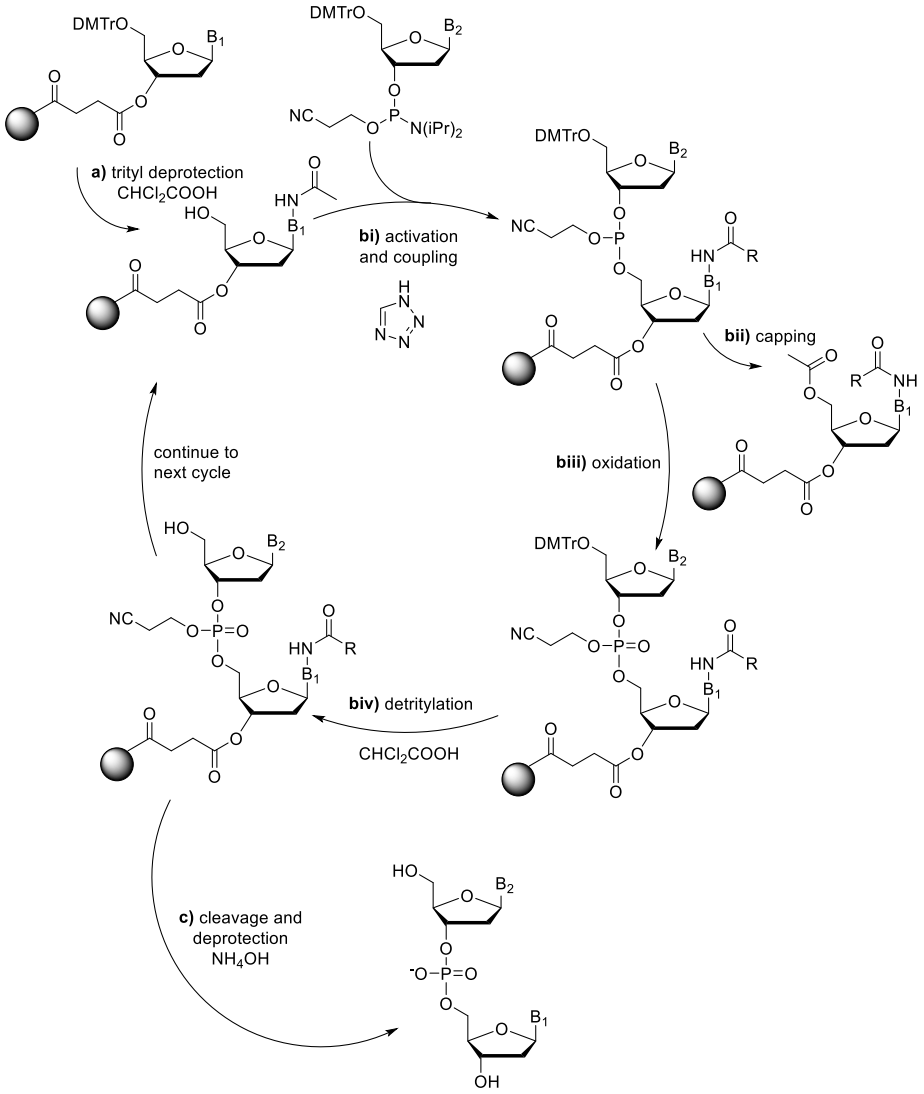
Oligonucleotide synthesis is common practice with many companies set up solely for the production of oligonucleotides. These oligonucleotides can range from small DNA primers only a few base pairs long, to sequences containing hundreds of nucleotides. As well as making simple oligonucleotides based on only the canonical bases, many manufacturers offer custom oligonucleotides which can be functionalised in a range of ways. After gene sequencing, oligonucleotides based on the target sequence can be synthesised to allow for detection.

Oligonucleotides are typically synthesised *via* the phosphoramidite method, allowing for the inexpensive synthesis of nucleic acids with defined sequences. Solid-phase synthesis is not limited to 2'-deoxynucleosides, as this method can be used for the synthesis of RNA, or modified nucleic acids such as locked, branched, or peptide nucleic acids.

In the phosphoramidite method, first developed in the 1980s by Caruthers *et al.*, the desired oligonucleotide is built from protected nucleoside monomers.[31] The 5'- hydroxyl position is protected with the acid-labile 4,4'-*O*-dimethoxytrityl (DMTr) protecting group to prevent polymerisation of the oligonucleotide, and the 3'-OH is activated through the use of nucleoside phosphoramidites. Solid phase synthesis of oligonucleotides is usually performed from the 3'-end to the 5'-end. The first stage of oligonucleotide synthesis is attaching the first base onto the solid-support, typically controlled pore glass (CPG), after which point the desired sequence can be built through a series of couplings, oxidations and washes (Scheme 1.5). The first nucleobase is subjected to trityl deprotection (a), through treatment of the solid-supported base with acid, (typically dichloroacetic acid) to yield the nucleophilic 3'-OH. The next base is activated through treatment of the phosphoramidite with a tetrazole catalyst, displacing DIPEA, creating a more electrophilic site for the 3'-OH to attack (bi).

The growing oligonucleotide is treated with acetic anhydride to cap any unreacted 5'-OH residues (bii), followed by oxidation of the phosphite to the phosphate with elemental

iodine (biii), and the nucleic acid is again subjected to the cycle until the desired sequence is achieved. The nucleic acid is then cleaved from the solid-support and any protecting groups present (*e.g*. amide protecting groups of exocyclic amines, or cyanoethylphosphate groups) are cleaved through treatment with concentrated ammonia (c). The crude oligonucleotides are then purified *via* HPLC to give the desired oligonucleotide either as a lyophilised powder or as an aqueous solution. Both are stable and can be stored at –20 °C for extended periods of time without decomposition.



Scheme 1.5 Oligonucleotide synthesis *via* the phosphoramidite method

Broadly speaking, there are four ways in which to label oligonucleotides. These are end, sugar, phosphate, or base labelling (Figure 1.4).

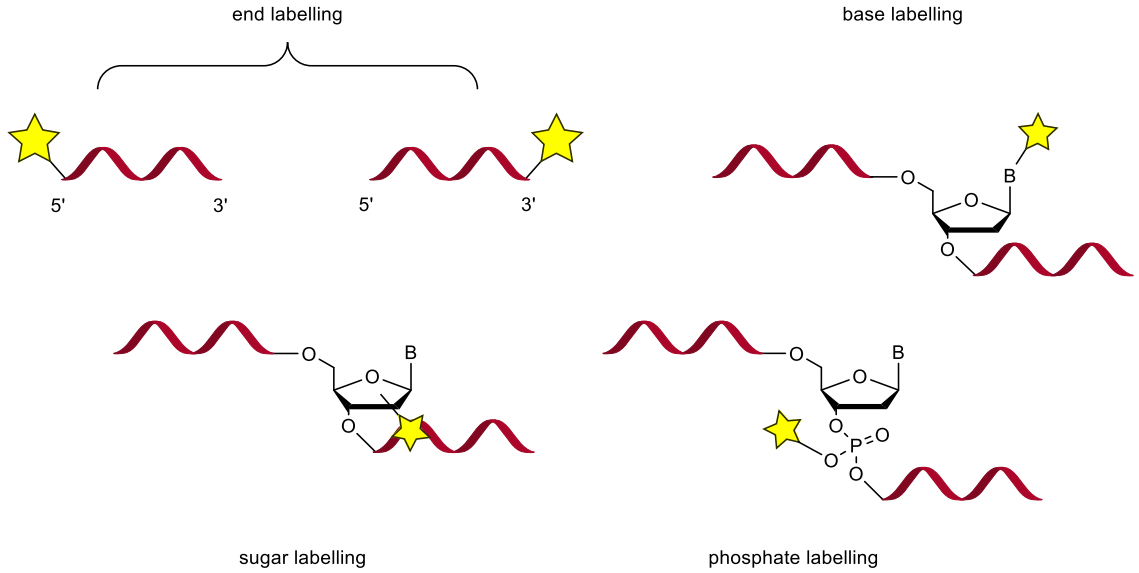
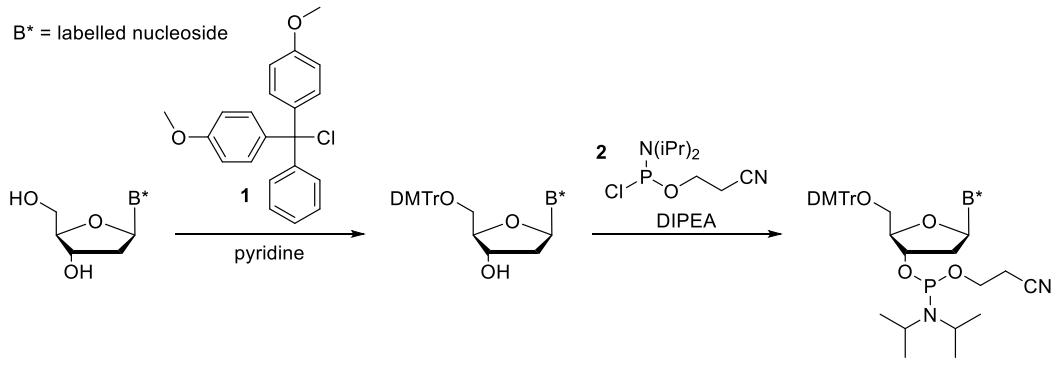


Figure 1.4 Modification sites of oligonucleotides

Synthesis of labelled units is generally conducted *via* the phosphoramidite method, which involves protection and activation of the label as the DMTr phosphoramidite (Scheme 1.6).[32] This is achieved through treatment of the nucleoside with 4,4'-dimethoxytrityl chloride (**1**) in pyridine to first yield the 5'-protected nucleoside, followed by activation of the 3'-OH residue with 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (**2**).



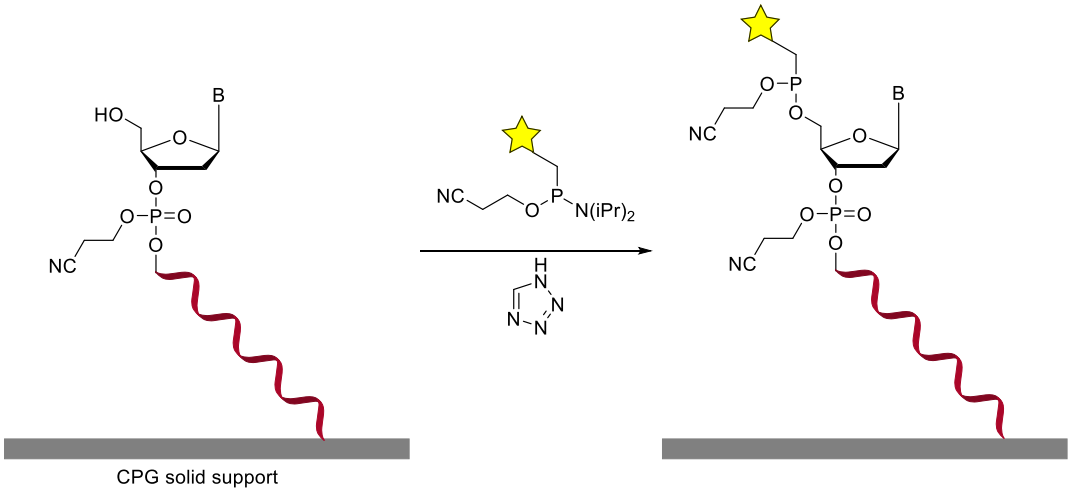
Scheme 1.6 General method for conversion of labelled nucleosides to the DMTr phosphoramidite

The location of introducing the label must be considered, as for the label to have potential as a successful probe, it must not impair the chemical or biological function of DNA. As such, labelling must not interfere too much with electronics to continue to allow for base pairing and π-π stacking, or change the shape of the unit to not obstruct enzymatic activity or α-helix formation. Non-canonical bases are more diverse in size and shape, however sacrifice base-pairing for improved fluorescent properties. External fluorophores or redox active

moieties allows for fine-tuning of the measured property, without confinements of base-pairing ability, which would significantly affect the properties of any label.

1.2.1. End-capping

End-capping is the simplest method in which to label oligonucleotides. Traditional oligonucleotide synthesis occurs in the 3'- to 5'- direction, and as such, when the final base is installed, the oligonucleotide can simply be treated with a phosphoramidite of the desired label (Scheme 1.7).[33] This caps the oligonucleotide, which is then subjected to oxidation to give the P(V) species, and this can then be cleaved from the support and purified *via* the normal process. 3'-labelling can also occur in this manner if the oligonucleotide is bound to the solid-support in the opposite direction.



Scheme 1.7 5'-end capping of oligonucleotides

While offering a reliable method to synthesise custom sequences, the incorporation of custom labels can be prohibitively expensive, as oligonucleotide manufacturers often require minimum quantity orders. In addition, the introduced label must not include any acidic or enolisable protons due to the treatment of the oligonucleotide with high concentrations of basic reagents, which would cause probe decomposition, and can limit the generality of the labels.

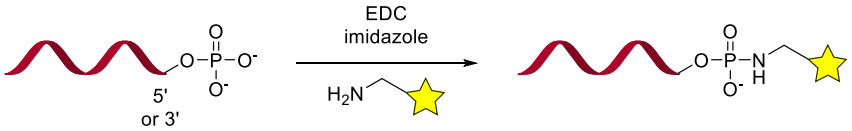
Labels can instead be incorporated enzymatically, with the introduction of 5'- and 3'- labels being possible with the use of ligases or kinases.[34] It is not always a practical solution, however, as setting up protocols for enzymatic incorporation of labels can be difficult and expensive, particularly if conducted on small scales. Depending on the final application of the probes they may also be required to undergo further purification using methods such as membrane filtration, spin column chromatography, or HPLC, further adding to costs.

3’-labelling can be achieved by treatment of the desired oligonucleotide sequence with labelled dNTPs and enzymes such as Klenow or T4 DNA polymerase.[35,36] T4 DNA polymerase synthesises DNA in the 5’-3’ direction and also has 3’-5’ exonuclease activity, which is exploited to allow for synthesis of labelled probes. The enzyme will cleave bases from the 3’-end, then replace with dNTPs containing the desired labels. Care must be taken to avoid long reactions because the enzyme will revert to exonuclease activity and degrade the newly labelled DNA after all dNTPs have been exhausted. T4 polymerase has found applications in radiolabelling, sequencing, and more recently, electrochemical detection.[37,38]

The enzyme poly(A) polymerase will add a single adenosine unit, which can then undergo cleavage to leave a ^{32}P labelled phosphate residue at the 3’- end for use as a radiolabel.[39] This can also be applied to RNA labelling with transferase and dNTPs, which only extend in the presence of 3’-OH, and as such can be used to incorporate a single labelled base.[40–42]

Modification at the 5’-end of DNA can also be achieved through enzymatic methods. T4 polynucleotide kinase can be employed which undergoes a forward and exchange reaction, adding ^{32}P to 5’-OH DNA. This methodology was used to allow T4 polynucleotide kinase to label substrates to give non-nucleosidic bases at the 5’-end.[43]

Chemical labelling can also be achieved with the diimide coupling agent *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride (Scheme 1.8).[44] This allows a wide range of labels to be incorporated and does not rely on potentially sensitive enzymatic conditions, however labelled oligonucleotides require purification after the reactions to ensure only desired probe is isolated.



Scheme 1.8 Carbodiimide coupling of phosphates

1.2.2. Phosphate modification

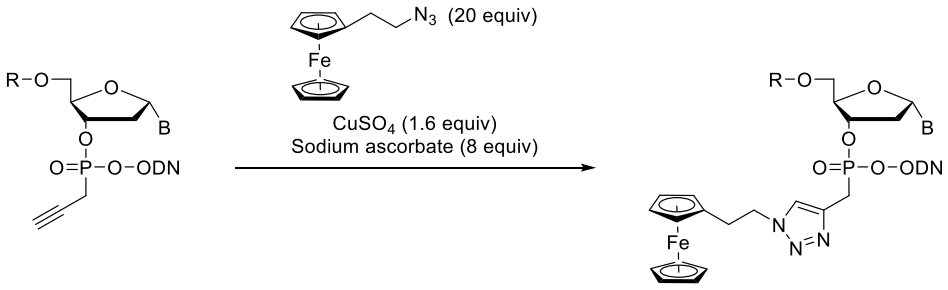
Radiolabelling was the first instance of introducing labels directly onto DNA probes and was achieved through phosphate modification. Other phosphate modifications have since been developed to introduce electrochemical and fluorescent labels directly onto the phosphate backbone. Labelling of phosphate linkages can lead to unpredictable and undesirable biological properties due to diastereomer formation.[45] In an unmodified oligonucleotide, the two non-bridging oxygens are resonance stabilised leading to equal negative charges, however introduction of substituents in the non-bridging position leads to the formation of a new stereogenic centre and a pair of diastereomers, which can be difficult to separate. The

unnaturally shaped diastereomer may lead to significant destabilisation in the duplex due to the change in angle of Watson-Crick base pairing and as such, phosphorous-modified oligonucleotides are generally more stable to nucleases, and have found applications in the modification of therapeutic agents.[46]

Phosphate modification of ODNs through the introduction of alkyne moieties allows for the highly regioselective copper-catalysed azide-alkyne cycloaddition (CuAAC) reaction to occur. This strategy was simplified by Krishna and Caruthers, who described the development of a new phosphoramidite building block with the replacement of one of the non-bridging oxygens with ethyne (Scheme 1.9).[47] This was then used to create ODNs, and the alkyne was functionalised during solid-phase synthesis *via* CuAAC chemistry, utilising the copper (II) sulfate and sodium ascorbate (NaAsc) catalyst system. It was shown that the T_m of duplexes formed were higher than those with standard linkers, and the duplex was significantly more stable to digestion by exonuclease enzymes.

Scheme 1.9 Ethynyl modified phosphates for phosphate modification

Duplex destabilisation from phosphate modification can be counteracted by using α-anomeric ribose. Chaix *et al.* described the use of the modified α-oligonucleotides (Scheme 1.10), allowing for substitution on the phosphate without causing destabilisation of the duplex.[48]

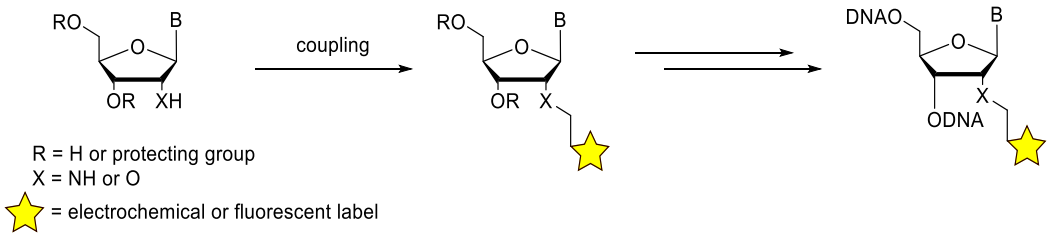


Scheme 1.10 α-oligonucleotides for improved dsDNA stability

1.2.3. Sugar modification

The ribose moiety is another position in which to label DNA. Incorporation of labels in the 2'-position of ribose has been used extensively to avoid substitution of the nucleobase. Modification of the ribose moiety is usually carried out on gram-scale, synthesising the 5'-

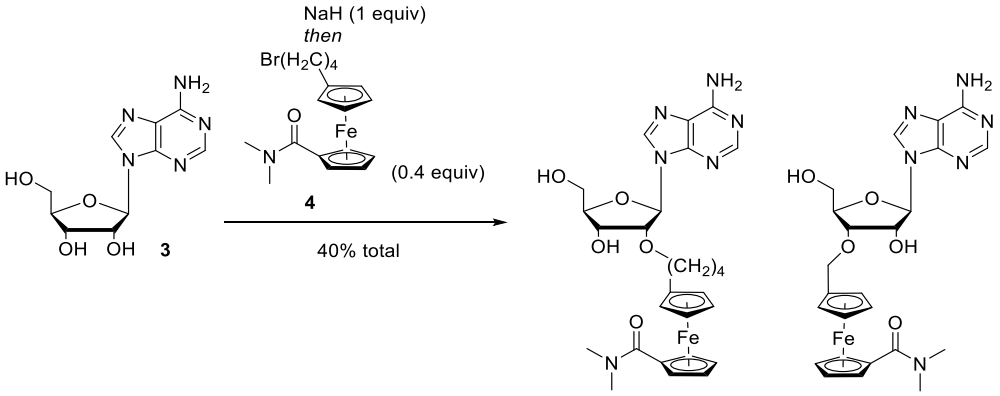
O-DMTr phosphoramidites and incorporating the modified units during solid-phase synthesis. This methodology has been followed to successfully introduce fluorescent and electrochemical labels.[49] Introduction of a label directly to the ribose moiety usually occurs in the 2'- position to allow for standard 5'- 3'- connectivity in an oligonucleotide (Scheme 1.11).



Scheme 1.11 General 2'-modification of ribonucleosides

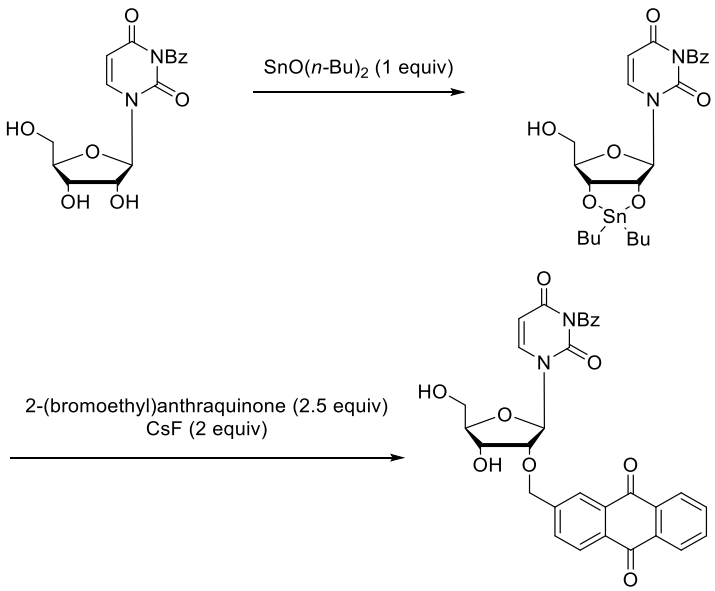
The reaction of unprotected ribose allows for deprotonation and subsequent treatment with an appropriate electrophile. Regioselectivity, however, is far from ideal and substitution reactions tend to give ratios of 2:1 in favour of the 2'-product.

The conjugation of ferrocenyl alkyl bromide **4** to adenosine **3** *via* nucleophilic substitution was achieved by Meade *et al.* (Scheme 1.12),[50] and although the reaction requires separation from the regioisomeric by-product, it offers a method to functionalise ribose without the need for further protecting group strategies.



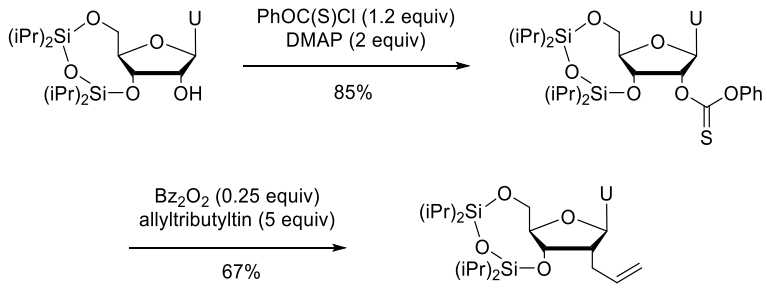
Scheme 1.12 2'-*O*- electrochemical substitution of ribose

Yamana *et al*. first introduced fluorescent labels into the ribose moiety of nucleosides.[51] These were incorporated into the 2'-position of ribose through substitution of dibutylstannylenes with alkyl bromides (Scheme 1.13). It was shown that the introduction of the anthraquinone (AQ) unit resulted in improved stability of the oligonucleotide upon duplex formation, which was attributed to the intercalation of the AQ unit between the base pairs.



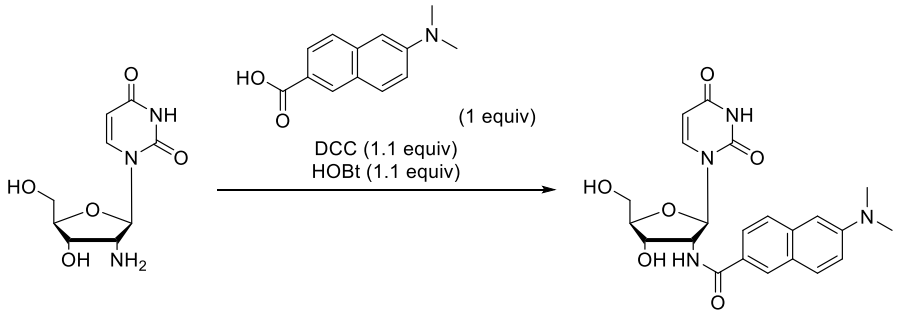
Scheme 1.13 Introduction of fluorophore *via* nucleophilic substitution

Simultaneous protection of the 5'- and 3'- hydroxyls was achieved with siloxane protecting groups (Scheme 1.14), which allowed Usman and co-workers to synthesise 2'-modified uridines for incorporation into ribozymes.[52] Ribozymes are often unstable in biological media due to the widespread existence of RNases. Modification at the 2'-position allowed for the synthesis of catalytically active RNA for use as ribozymes, with increased resistance to nuclease activity. This was achieved through the substitution of 2'-hydroxyls *via* deoxygenative allylation reactions.



Scheme 1.14 5',3'-*O* protecting strategy

Amide couplings remove regioselectivity issues if there is an amine residue in the 2'-position instead of the hydroxyl. Amide couplings have been used to introduce fluorophores in the 2'-position (Scheme 1.15),[53] as well as the use of sulfonyl chlorides as coupling agents.[54]



Scheme 1.15 Sugar labelling through amide coupling

1.2.4. Non-nucleosidic units

Non-nucleosidic units have also gathered much attention, which can either be achieved through replacing a single nucleotide such that there is a gap in the bases when oligonucleotides are synthesised, or by replacing the ribose with a label, onto which the base is conjugated (Figure 1.5).

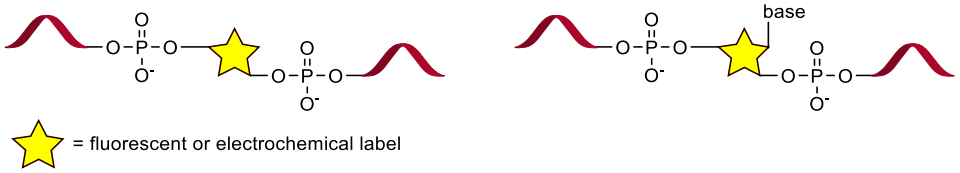
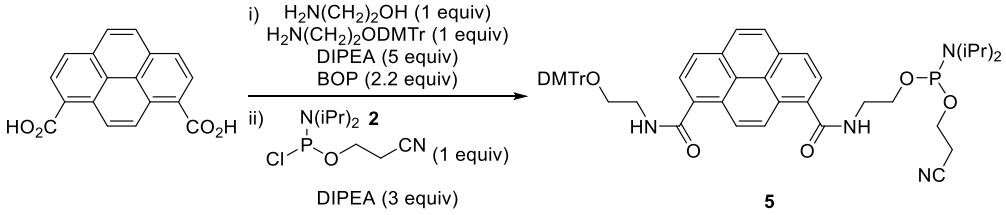


Figure 1.5 Non-nucleosidic labelling strategies

Pyrene has been extensively studied as a nucleoside mimic, as it readily forms fluorescent excimers. Langenegger and Häner developed non-nucleosidic pyrene units,[55,56] where the pyrene was introduced as a non-nucleosidic unit, replacing both the base and the ribose (Scheme 1.16) *via* the synthesis of non-nucleosidic DMTr phosphoramidite **5**. The single-stranded DNA had a single fluorescence emission, however upon duplex formation and intramolecular π-π stacking with a second pyrenyl unit, the pyrenes exhibited a second fluorescence emission.



Scheme 1.16 Interstrand stacking of non-nucleosidic pyrene units

Serinol has been employed as a ribose mimic.[57] The 1,3-diol mimics the 5',3'- hydroxyl linkage in ribose and can be used to enable substitution of a single nucleobase. The units

are synthesised *via* amide couplings to install the serinol unit, which are then protected and activated using standard DMTr phosphoramidite chemistry to incorporate in solid-phase ODN synthesis. Around the same time, Asanuma *et al.*[58] and Tucker *et al.*[59] developed serinol-modified DMTr phosphoramidites **6** and **7** for incorporation into oligonucleotides to detect for deletion mutations and SNPs (Figure 1.6). The introduction of DNA intercalators allowed for the detection of mutations *via* structural changes induced in the oligonucleotide, which cause changes in the emission spectra, attributed to a change in base pairing of the oligonucleotides.

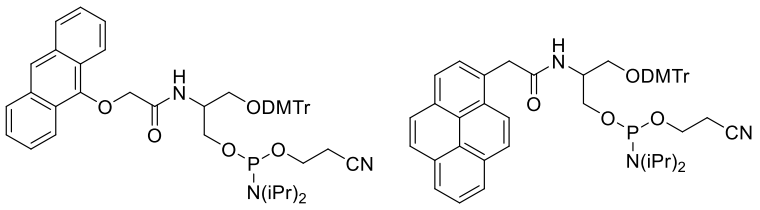


Figure 1.6 Non-nucleosidic fluorescent phosphoramidites

Non-nucleosidic units have also been studied *via* the replacement of the ribose moiety with redox active units. Tucker *et al*. reported the use of ferrocenyl thymidine as a replacement for two nucleobases.[60,61] Ferrocene thymidine **8** was incorporated into automated nucleic acid synthesis, coupling eight units together. The single-stranded ferrocene nucleic acid was shown to have UV absorption as expected and a single, reversible oxidation at 212 mV (CV, *vs.* Ag/AgCl). Duplex formation of the strand with complementary sequences was not studied.

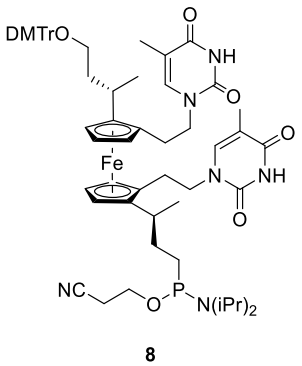


Figure 1.7 Ferrocenyl-thymidine

1.2.5. Base labelling

Much research has been carried out on the synthesis of modified nucleobases. Canonical bases are not inherently fluorescent, and as such the introduction of labels allows for

discrimination and detection of sequences. Generally, the introduction of fluorescent labels onto canonical bases involves ring fusion or substituent addition (Figure 1.8)[62] and through the introduction of extended π-systems the inactive bases can be made to fluoresce. The purine skeleton can be influenced readily, with the introduction of a single aromatic ring allowing fluorescence. Due to their comparatively smaller size, pyrimidine units tend to require a higher level of substitution to create fluorophores.

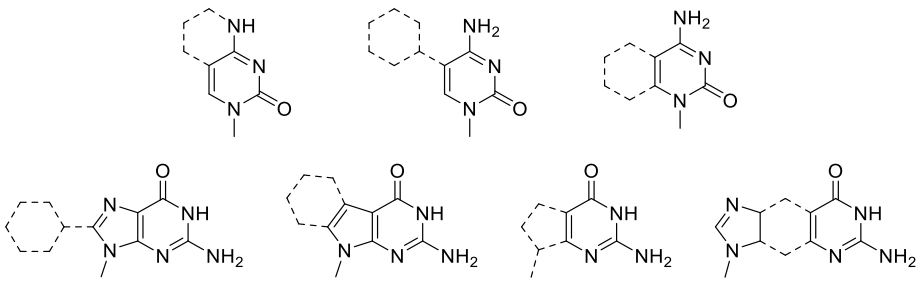


Figure 1.8 Possible modification sites of cytidine and guanine

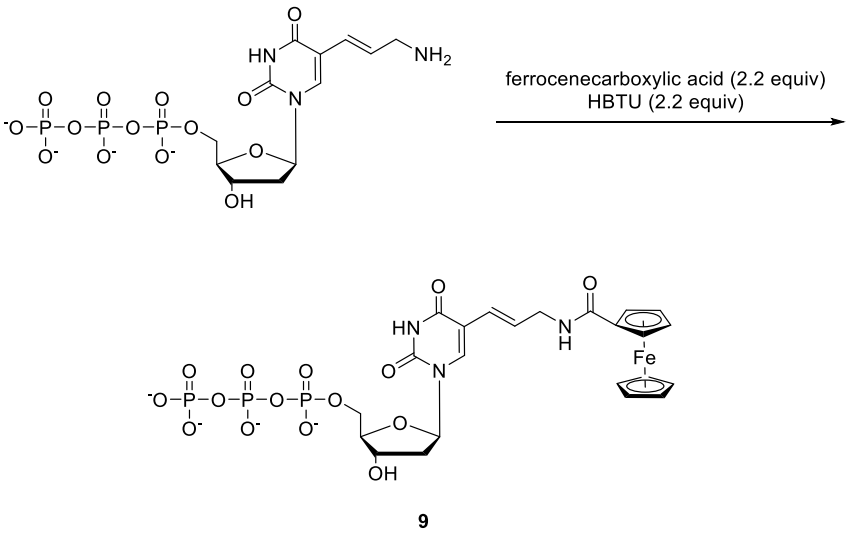
Uridine is the most commonly chosen nucleoside to attach an external label. The pyrimidine bases are easier to handle due to their reduced polarity compared with the purine bases. This allows easy synthesis of 5-halogeno pyrimidines, which can then be subjected to cross-coupling reactions to introduce the desired label.

After incorporation of the desired label onto a nucleobase, three main methods exist to introduce the nucleoside into a nucleic acid strand, namely direct, post-synthetic modification, and enzymatic incorporation. Each have their advantages and disadvantages associated however suitability is highly dependent on the end-user and scale at which the labelled DNA is required to be synthesised.

Enzymatic incorporation is the simplest, whereby labelled nucleosides are incorporated *via* modified dNTPs. This cheap method allows for the synthesis of long DNA strands, allows incorporation of many labelled units, and can be carried out in most laboratories with standard biology equipment. It does, however, limit the type of label that can be incorporated as they must not inhibit the enzyme's activity. The DNA obtained after enzymatic synthesis is difficult to use as DNA probes, due to the requirement for separation. As such, these are typically used as signalling probes, for example in nick translation reactions, or PCR, to indicate level of completion of the reaction.

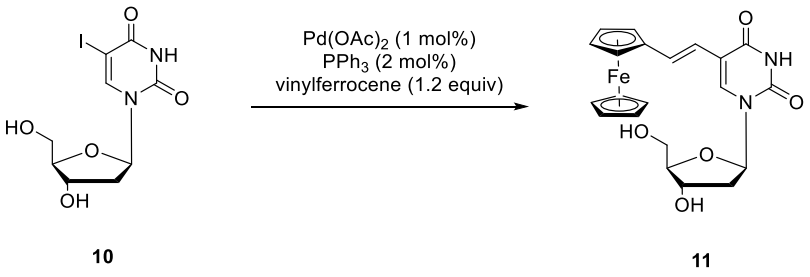
Modified dUTPs have been used to introduce ferrocene onto the base stack of DNA *via* enzymatic incorporation (Scheme 1.17).[63] The ferrocene unit was introduced through coupling of ferrocenyl acids and amine functionalised nucleoside triphosphates activated by HBTU. The labelled dUTP **9** was shown to have been successfully incorporated into DNA by the enzymes T4 polymerase and Klenow fragment, with data indicating a minor

destabilisation of the duplex and reduction in T_m. **9** was also shown to be a suitable substitute for dTTP in PCR, allowing for the synthesis of long DNA fragments.



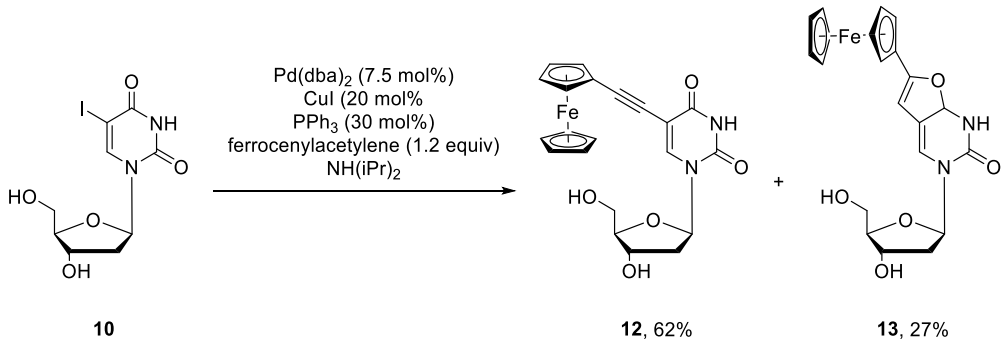
Scheme 1.17 Synthesis of ferrocenyl dUTPs

The conjugation of ferrocene onto nucleotides has been studied by several groups. Having a conjugated π-system allows for the charge transfer between the base and the ferrocene, therefore modifying its redox properties.[64] This label was synthesised from 5-iodo-2'-deoxyuridine **10** utilising the Heck reaction to yield the desired nucleoside **11**, which was then incorporated into the desired oligonucleotide sequence *via* the phosphoramidite method according to standard procedures.



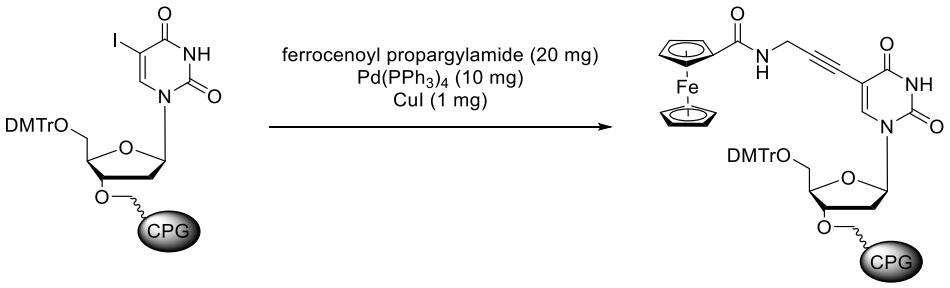
Scheme 1.18 Heck reaction between ferrocene and 5-iodo-2'-deoxyuridine

The conjugation of ethynylferrocene directly to uridine was also attempted,[64] however concurrent with previous studies by Yu *et al.*,[65] the Sonogashira cross-coupling does not proceed cleanly, instead yielding a mixture of the desired product **12** and the cyclisation by-product **13**. Under basic conditions, the iminolate of uridine forms slowly, which cyclises with the activated alkyne (Scheme 1.19). The yield of the cyclised product increases with increased reaction time.



Scheme 1.19 Undesired cyclisation of ferrocene and uridine under basic conditions

Cross-couplings of ferrocene and nucleosides have also been carried out "on-column", such that 5-iodo-2'-deoxyuridine is incorporated as a modified dT unit.[66] The DMTr phosphoramidite of 5-iodo-2'-deoxyuridine was incorporated into an oligonucleotide sequence, and the solid-phase synthesis was temporarily paused. At this point the resin was subject to Sonogashira cross-coupling, yielding the redox active nucleobase (Scheme 1.20). The reagents and catalysts for the Sonogashira reaction were washed away, and oligonucleotide synthesis was continued according to the standard coupling/washing procedure.



Scheme 1.20 "On-column" ferrocene cross-couplings

1.2.6. Post-synthetic modification

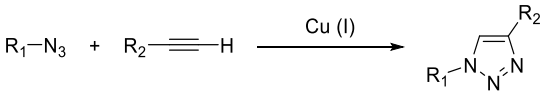
Post-synthetic modification is achieved by the introduction of a functional handle during oligonucleotide synthesis. This sequence is constructed and purified in the usual method, and after the desired sequence has been made and cleaved from the resin, the nucleoside is functionalised with the label of choice. Although more expensive than enzymatic incorporation, it allows site-specific introduction of labels, as the location of functional handles can be controlled more. Additionally, it allows possibly the widest range of functional group tolerance of the three methods, as the ODN is not subject to further removal from resins, or deprotection of various protecting groups required during solid-phase synthesis. The major drawback of this method is the significant purification challenge posed, as incomplete functionalisation, particularly in cases where multiple functional

handles exist, would result in complex mixtures of products with very similar solubility properties.

The first enzymatic methods used was nick translation, which was used to incorporate radiolabelled dNTPs.[67] In this method, the enzyme locates the nick (a gap in the dsDNA duplex) and extends the nucleoside while simultaneously digesting the complementary strand, therefore incorporating the labelled dNTPs. Other enzymatic methods include using modified dNTPs during PCR, however enzymatic methods yield mixed products which have target DNA already present and as such cannot easily be used to detect DNA without further purification.

"Click" chemistry was defined by Sharpless as any reaction that is high yielding, stereospecific, produces minimal waste, does not require chromatography, and can be carried out under mild conditions (*e.g*. low temperatures, low-boiling solvents).[68] The regioselective, copper-catalysed version of Huisgen's azide-alkyne cycloaddition,[69] was discovered simultaneously by Meldal *et* al. and Sharpless *et al.* and is the most well-known "click" reaction (Scheme 1.21).[70,71] Its applications span chemical and biochemical synthesis, allowing unprecedented access to a range of materials in highly specific, high yielding reactions.

The CuAAC is a strong candidate for post-synthetic modification, as it is biocompatible and can be carried out in aqueous solutions, thus allowing the oligonucleotides to be suspended in buffers and introducing the label at a later point. Click chemistry has been widely utilised in DNA modification due to its lack of reactivity with functional groups naturally found in DNA. The CuAAC has been used to functionalise nucleic acid oligomers, employed as a DNA crosslinking agent, supramolecular structure formation tool, phosphate surrogate, and as a surface immobilisation technique.[72]



Scheme 1.21 Copper catalysed azide-alkyne cycloaddition

The CuAAC of alkyne-functionalised oligonucleotides has been used to introduce several labels on the DNA backbone. The introduction of a number of alkynes on the backbone of DNA allows access to highly functionalised oligomers *via* post-synthetic modification (Figure 1.9).[73] Treatment of ODNs with six consecutive alkyne moieties afforded functionalised products in quantitative yields. This allowed for the conjugation of glucose, coumarin, and fluorescein azides onto the oligonucleotide. Comparison of flexible and inflexible alkyne moieties showed that using the flexible alkyne **15** allowed for quantitative

incorporation of azide labels such as glucose, biotin, and coumarin, however when using rigid structure **14**, the CuAAC reaction was inhibited, resulting in incomplete conversion to desired product. Two sequences were used as PCR primers, each containing two free alkynes sites. These primers allowed for the synthesis of PCR products up to 2000 bp long, giving access to clickable PCR amplicons.

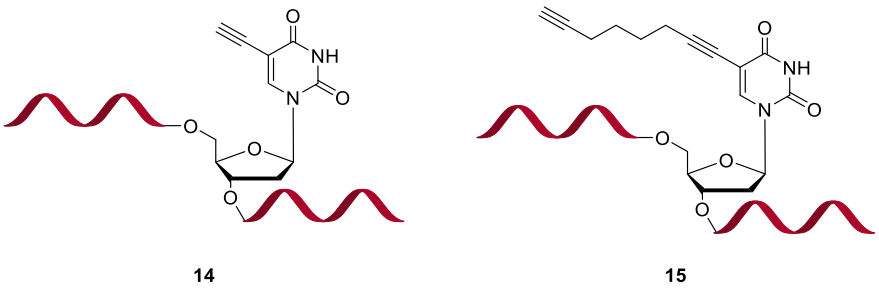
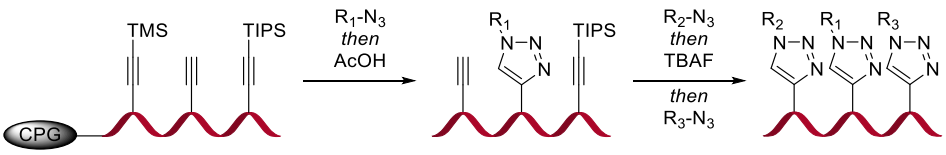


Figure 1.9 Alkyne functional handles on 2'-deoxyuridine

DNA CuAAC chemistry can also be achieved by introducing different protecting groups, so that multiple labels can be introduced onto a single strand post-synthetically (Scheme 1.22). The modular deprotection of alkynylsilanes can be exploited to introduce different labels in a site-specific manner when multiple functional handles are present. Carell *et al*. developed their previous DNA CuAAC strategy and used silyl protecting groups to modify the reactivity of alkynyl ODNs, allowing for the introduction of a range of fluorescent and biologically relevant labels.[74] By using this methodology, the desired label could be incorporated regioselectively, thus yielding the desired labelled sequence.



Scheme 1.22 Sequential CuAAC reactions on oligonucleotides

The CuAAC has proven itself as a useful tool for functionalising DNA, pre- and post-oligonucleotide synthesis. It can be utilised to introduce labels on the sugar, phosphate, and base, as well as being used to functionalise other nucleic acids such as PNA.

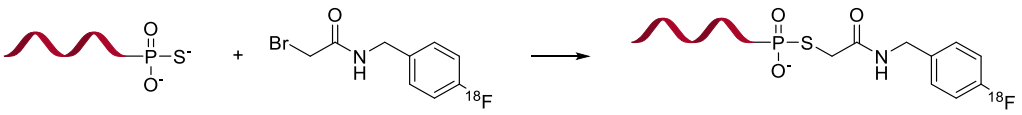
1.3. Nucleic Acid Detection Methods

The development of a range of labelling techniques has allowed for the design of different detection methods. Dependent on the type of label used, the applications can vary widely, with each detection method having its associated benefits and drawbacks. Further development has allowed the applications to move from the laboratory benchtop, to field-based point-of-care (POC) diagnostics.

1.3.1. Radiolabelling

Labelling of DNA was first achieved through the introduction of radioisotopic labels to isolate complementary RNA and DNA.[75] These were detected through hybridisation with agar DNA columns and the level of radioactivity was measured to calculate the amount of radioisotope incorporated. Incorporation of labelled fragments is usually achieved by enzymatically introducing ^{32}P labelled triphosphates at the 5'- and 3'- ends.[67,76] Labelled nucleic acids are typically detected through radiography film, positron emission tomography, or *via* Cherenkov radiation.

Radiolabelling of compounds has drawn much interest for *in vivo* imaging, whereby incorporating β- or γ- emitters onto an oligonucleotide allows for imaging *via* positron emission tomography to assess cell uptake of antisense oligonucleotide therapeutics (Scheme 1.23).[77,78]



Scheme 1.23 Radiolabelling of oligonucleotides

1.3.2. Colourimetric and fluorescent detection

Colourimetric biosensors are easy to interpret and do not require expensive machinery.[79] The ease of this detection method is exemplified by home pregnancy tests. A simple blue strip indicates pregnancy, which is unequivocally interpreted by the user. Colourimetric DNA sensors, however, have not achieved such success. Chee *et al*. developed colourimetric DNA arrays to analyse gene expressions.[80] The probes were labelled with a red protein, and the target with green fluorescein, and the red and green light intensity was compared upon hybridisation of the oligonucleotides. A decrease in intensity indicates the location of a SNP, and further decreases are seen when two polymorphisms occur within proximity to one another.

Colourimetric detection has not found widespread implementation in DNA diagnostics due to its limited sensitivity, however fluorescent molecules have allowed for the development of a huge range of detection methods. These range from non-specific interactions with any dsDNA, to highly sensitive probes for single nucleotide polymorphism detections. The high sensitivity of fluorescent-based detection methods allows for the detection limits down to a single molecule, allowing for early detection when pathogenic material is low, or for detection of highly toxic substances at low concentrations.

A fluorophore is a compound that absorbs and re-emits light when irradiated at the correct wavelength of light. Such compounds typically contain conjugated aromatic rings or extended π-systems. The excitation of an electron to a higher energy state and subsequent relaxation allows the electron to emit the energy in the form of light, if the excitation beam and the HOMO-LUMO gap are sufficiently close in energy. The absorption and emission spectra can be analysed and the emission intensity is directly related to the concentration of fluorophore present according to the Beer-Lambert law, therefore fluorescence is an excellent choice for quantitative techniques.[81] Organic compounds that emit coloured light can be used for many applications such as dyes and solar cells. High molar extinction coefficients of highly conjugated aromatic compounds allow the dyes to be used in low concentration, whilst yielding high signals.

DNA intercalators are small, usually highly conjugated, aromatic compounds, which insert between the base pairs in the double stranded form of a nucleic acid (Figure 1.10). This is mainly due to π-π stacking, an attractive force involved in chemical and biological recognition events, as well as DNA binding and protein structure.[82] Upon intercalation, the rigidity of the intercalator is increased, thus extending its π-system allowing an increase in quantum yield of the fluorescence.

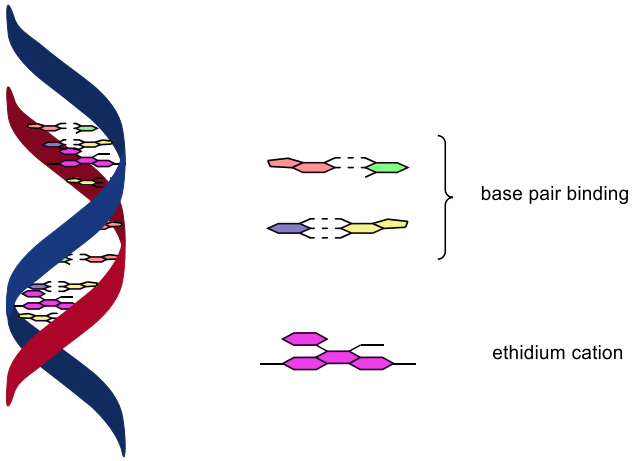


Figure 1.10 Ethidium bromide DNA intercalation

Intercalators were first identified as a tool for visualising DNA in the 1960s.[83] It was discovered that upon treatment of DNA in high concentrations of salt, ethidium bromide (EB) forms a fluorescent complex with the DNA. In the absence of DNA, ethidium bromide is mildly fluorescent, however introduction of dsDNA or RNA results in a large increase in the quantum yield, and a change in the absorption spectrum. There is clear interaction between EB and nucleic acids, as shown by the isosbestic point upon mixing of EB with DNA or RNA. It was also found that EB and DNA form a complex in the single-stranded state, but this was attributed to the ionic attraction between the cationic EB core and the negatively charged phosphates. As a result, fluorescent detection of DNA while using intercalators gives relatively high background signals due to the ssDNA interactions.

Intercalators have been used extensively in gel electrophoresis, where the DNA is separated based on charge and size. Since the introduction of EB into DNA detection, other intercalating dyes have been developed which have improved fluorescence properties over EB, which are largely designed around EB's core phenanthridine structure, as well as acridine, and cyanine based dyes.[84] Gel electrophoresis, however, has inherent drawbacks due to the length of time required to achieve results. Gels are referenced against DNA of known length, and as such the readout gives a value of size alone. The implications, therefore, are limited specificity, with little possibility of multiplexing amplification tests to detect for multiple targets simultaneously.

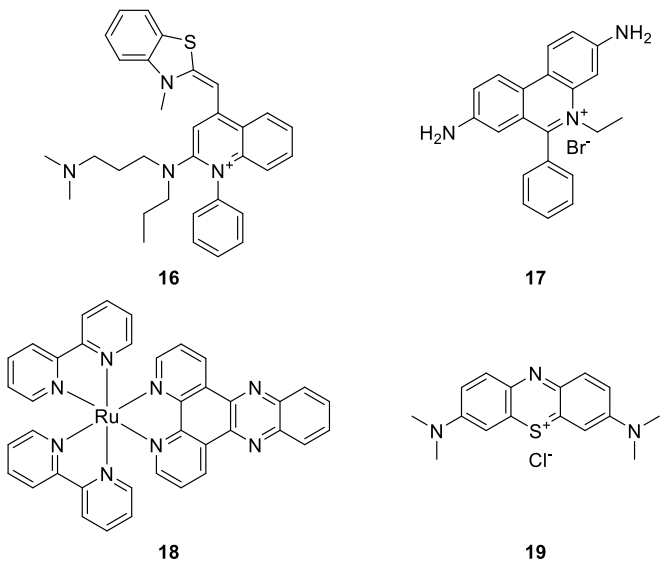


Figure 1.11 Structures of common DNA intercalators (Top left, clockwise, SYBR green, ethidium bromide, $Ru(bpy)_2(dppz)$, methylene blue)

EB, **17**, and SYBR green, **16**, are dyes that bind to dsDNA. Both are common biological reagents that are used to stain DNA. For example, tracking the progress of PCR assays in the absence of any labels is achieved through gel electrophoresis. Rasmussen *et al*. used SYBR

green to quantify the progress of PCR. By monitoring the fluorescence during melting, the double stranded DNA is denatured, dramatically reducing the fluorescence.[85] The temperature at which this occurs, and the amount the fluorescence decreases by can then be attributed to the amplification product and the user can then quantify the amount present.

As well as organic fluorophores, organometallic complexes have been used to indicate the presence of dsDNA. Hartshorn and Barton used ruthenium bipyridyl/dipyridophenazine (**18**) complexes to indicate the presence of dsDNA.[86] In aqueous solutions the complex does not photoluminesce. However, upon introduction of dsDNA, the probe's luminescence increases by a factor of 10^4 as the metal complex intercalates to dsDNA and turns on the luminescence.

Asanuma *et al.*, used internally labelled oligonucleotides, where the insertion of a pyrene phosphoramidite on both sides of a base was used to discriminate between wild type and deletion mutagenic DNA.[58] In the presence of the wild type, the probe and target hybridise fully and the pyrene intercalates, thus reducing the fluorescence to very low intensities (Figure 1.12). In the target with a single nucleotide deletion, the probe has a 3-unit mismatch, which sits outside the duplex, increasing the fluorescence by a significant amount. The deletion portion was increased to two base pairs and the system still shows a significant increase in signal upon hybridisation with mutagenic DNA.

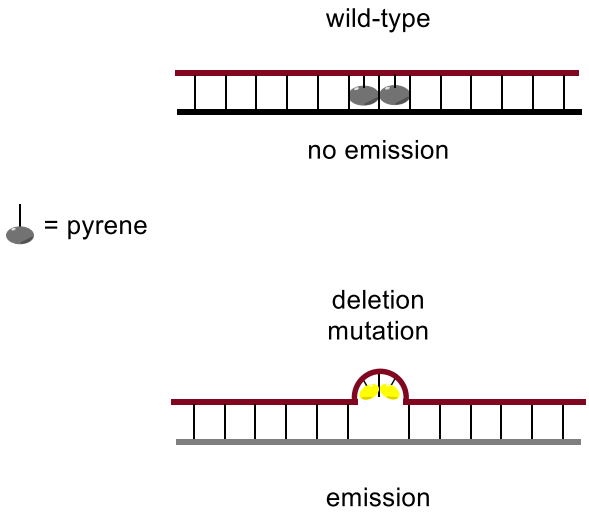
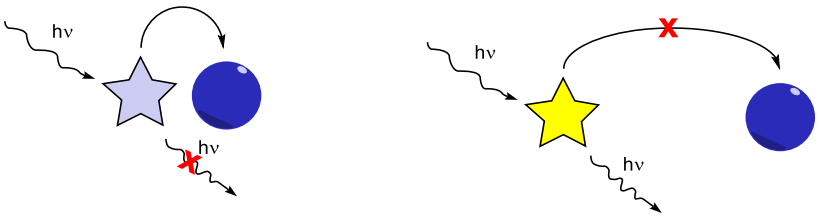


Figure 1.12 Detection of deletion mutations by fluorescence

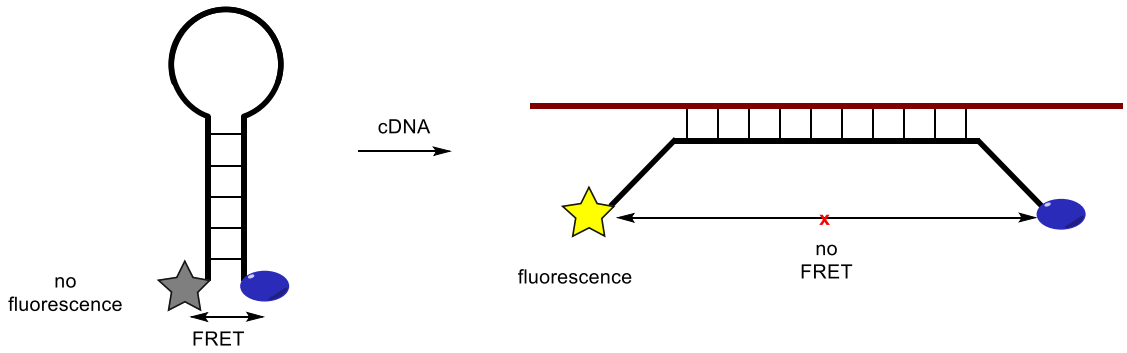
DNA intercalators suffer from high background signals, as ssDNA hybridisation is still a major contribution to overall signal. As such, probes in which the labels are covalently attached are more favourable.

To create a switch, a quencher molecule is sometimes added to turn off the fluorescence. These molecules are chosen so the energy of the HOMO-LUMO gap of the quencher matches that of the fluorescence. If the molecules are sufficiently close to each other, in both space and energy, Förster resonance energy transfer (FRET) will occur, and the quencher will relax to the ground state, releasing the energy as heat, thus no (or very little) fluorescence happens (Scheme 1.24). When the distance between the fluorophore and quencher is increased, FRET can no longer happen and the energy will be released as light.



Scheme 1.24 Förster resonance energy transfer

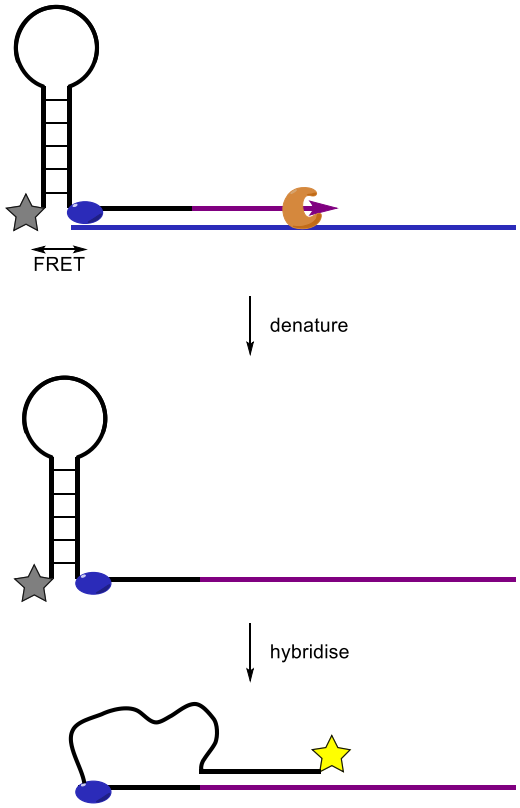
Folding-based biosensors are another method which can be used to detect pathogenic DNA. Molecular beacons were first developed by Tyagi and Kramer, exploiting the highly specific nature of DNA binding.[87] As with many other biomolecular recognition events, the binding of complementary DNA (cDNA) does not result in easily measurable changes of its physicochemical properties. The introduction of a label onto oligonucleotide probes allows for the detection of the binding event. Molecular beacon probes are based on hairpin structures, whereby a single oligonucleotide is labelled at opposite ends with a fluorophore and quencher, relying on proximity of the labels to each other (Scheme 1.25).



Scheme 1.25 Molecular beacons

These probes consist of a loop of DNA capped in the 5'- and 3'- ends with a naphthalene fluorophore and an azobenzene quencher. The probe has a self-complementary section at both ends which force the formation of the hairpin in the absence of cDNA, bringing the fluorophore and quencher in proximity to each other. However, upon introduction of cDNA, the probe unfolds and binds to the target DNA, allowing the naphthalene to fluoresce freely.

Whitcombe *et al.* described the use of self-probing amplicons that fluoresce upon amplification of the target sequence.[88] In this example, the primers used to indicate the start of the PCR product were also the probes, labelled with fluorescein at the 5'- end and a methyl red quencher on the oligonucleotide 30 base pairs away (Scheme 1.26). The primers also included a hexaethylene glycol monomer adjacent to the methyl red, which inhibits the action of the polymerase to ensure the extension products are not copied. The first PCR cycle extends the primer, and the product is subjected to another round of denaturing. The target now contains a complementary sequence to itself, so upon cooling the probe folds and anneals to the target sequence. This increases the distance between the fluorophore and quencher (*cf.* molecular beacons), allowing the real-time production of signal to be detected.



Scheme 1.26 Self-probing amplicons

Livak *et al.* exploited the exonuclease action of *Taq* polymerase to digest probes with fluorophores and quenchers attached.[89] Upon extension of the primer, the probe is digested, releasing fluorescein into solution, leaving a sufficient distance between the quencher to allow fluorescence to occur. The group investigated the effect of distance between the fluorophore and quencher, using distances of as low as 2, and as large as 26 base pairs. Even with a distance of 26 base pairs between them, the fluorescence was still quenched. Upon hybridisation (and digestion), the intensity of the fluorescence increased dramatically. However, when the fluorophore and quencher are only two base pairs away from each other, the enzyme does not release a small enough fragment to allow the quenching to stop.

Many organic molecules exist that have a range of absorption and emission spectra, which allows for multiplex capability for the detection of several targets in a single test. Lee *et al.* used two fluorescein compounds to discriminate between two alleles of the gene responsible for cystic fibrosis.[90] The labels, with a difference in λ_{max} of roughly 20 nm, were attached to an oligonucleotide with quenchers. Both probes were then subject to PCR with mutagen, wild-type, both mutagen and wild type, and blanks, demonstrating that the two can be differentiated in the same solution.

1.3.3. Electrochemical detection

Application of fluorescent DNA detection is usually confined to research laboratories due to the expensive nature of the machinery, although recently more devices are being integrated into point-of-care settings. Although many examples of detection *via* fluorescence are common in the literature and fluorescent detection is more widespread, electrochemical detection methods possess advantages over optical methods, and as such these methods are still of great interest to further develop.

Electrochemical detection holds many advantages over their fluorescent counterparts. Fluorescence uses bulky, precise devices that require calibration, often including lasers, meaning the cost of the devices is often extremely high and the equipment is difficult to transport due to the need for calibration and the sensitivity of the machine.[91] Conversely, electrochemical detection can be made significantly cheaper given the robustness of the equipment. No lasers are required, immediately reducing the cost, and the machines can be made from cheap materials – metal wiring, circuit boards, and computer chips, all of which are mass produced and as such a potentiostat can now be custom built for a low cost.

Electrodes are easily screen-printed and potentiostats can be miniaturised, allowing rapid point-of-care detection of pathogens to be an affordable technology. Analysis of samples is quick, taking only a number of seconds and can be conducted on turbid samples, reducing the need for purification. Detection of DNA can be achieved with or without the use of amplification techniques such as PCR.

The last three decades has seen a dramatic increase in publications in the field of electrochemically-labelled nucleic acids (Figure 1.13), leading to the development of a number of techniques now available for reliable, sensitive, and rapid detection of nucleic acids in both research and clinical settings.

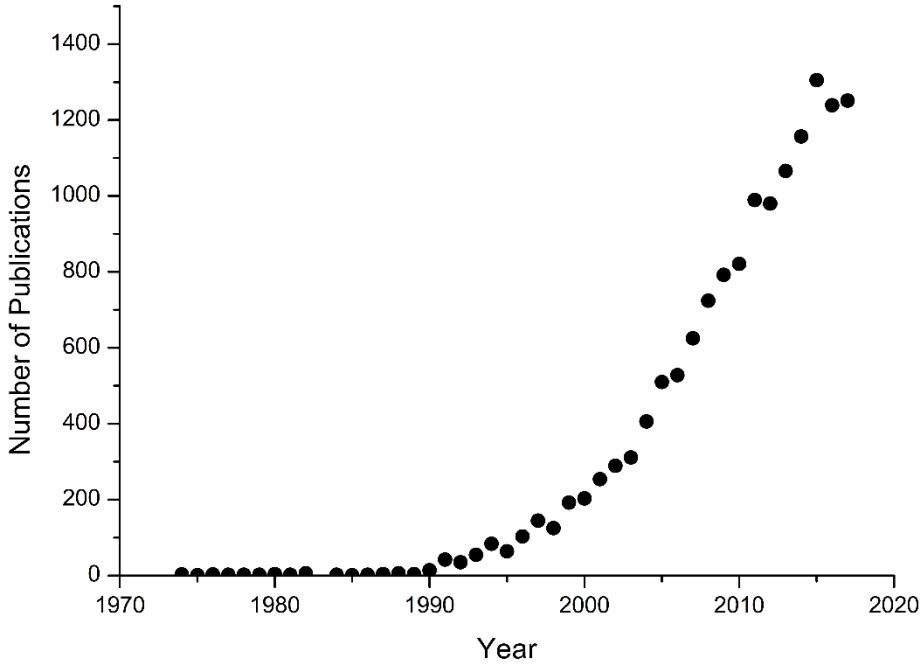
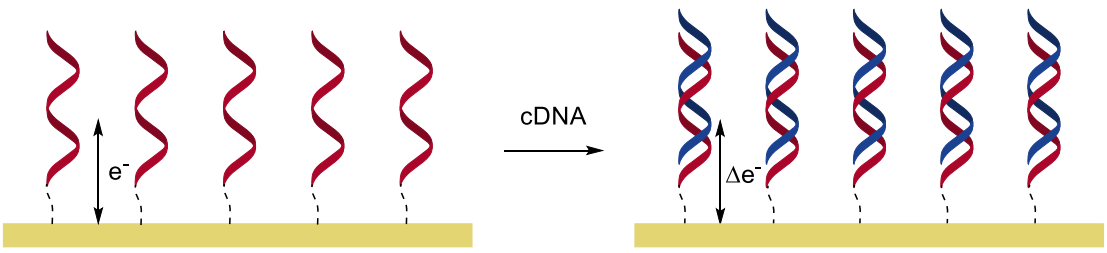


Figure 1.13 Publications per year on electrochemistry of nucleic acids. Numbers taken from Web of Science query for (electrochem* AND DNA) OR (electrochem* AND nucleic acid*).

1.3.3.1. Surface-immobilised detection methods

Electrochemical detection of DNA is typically achieved through adsorption onto an electrode surface and the presence of target DNA will elicit a structural or electronic change in the probe, therefore leading to a change in signal (Scheme 1.27). To create electrochemical sensors, the nucleic acid is first synthesised, and immobilised on an electrode, which are typically made from carbon, indium tin oxide (ITO), or gold. The electrode surface is washed to remove any non-specifically bound DNA from the surface, ensuring that only the desired, correctly bound probes remain.

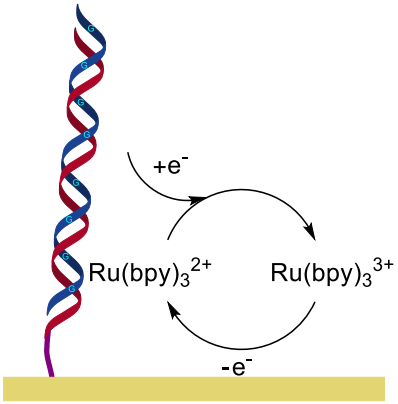


Scheme 1.27 Typical electrode surface immobilised DNA detection

There is a sensitivity battle to be considered when using surface-immobilised techniques. An increase in probe density allows for a higher signal due to increased number of redox

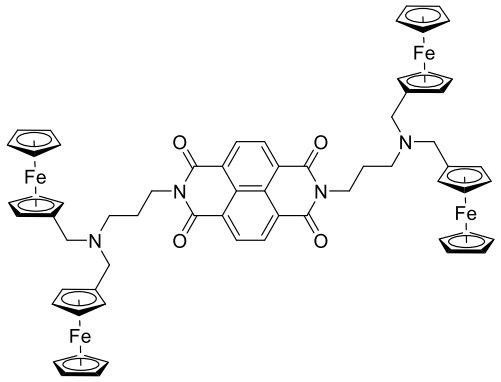
active sites available, however increasing the number of probes on the surface of the electrode too far also results in a decrease in sensitivity. Too high a probe density on the electrode surface would inhibit hybridisation events, as it would limit the amount of target DNA that can fit between the probe strands. As such, blocking layers such as mercaptohexanol are deposited simultaneously to ensure that the correct spacing of DNA is achieved. The adsorption of thiol-capped DNA to a gold surface can be controlled by changing the concentration of DNA, and varying the buffer and mercaptohexanol concentration, allowing uniformity of the DNA layers deposited.[92]

Label-free amperometric detection of nucleic acids can be achieved *via* the use of guanine-free probes (Scheme 1.28).[93] Guanine can be replaced with inosine, a naturally occurring nucleoside which still binds to C residues but has different E_{ox}, therefore inosine in the probe and any G residues in the target can be discriminated. Detection either occurs through direct oxidation from the probe on a carbon electrode, or *via* the redox cycling of $Ru(bpy)_3Cl_2$ using ITO electrodes. The number of guanosine residues shows a linear relationship with current, and as such the sequence and number of repeat units can be determined.



Scheme 1.28 Ruthenium oxidation of guanosine residues

Label-free detection of DNA can also be achieved through redox-active DNA intercalators (Figure 1.14).[94,95] DNA intercalators have been used in electrochemical detection of DNA, however, concurrent with fluorescent based intercalators, discrimination between ssDNA and dsDNA is not perfect and as such high backgrounds are often observed. Takenaka *et al*. developed the diimide derivative **20** to detect for DNA methylations in PCR amplicons. The capture probes were immobilised on gold electrodes and the current upon hybridisation can be interpreted to determine whether wild-type or mutant DNA is present.



20

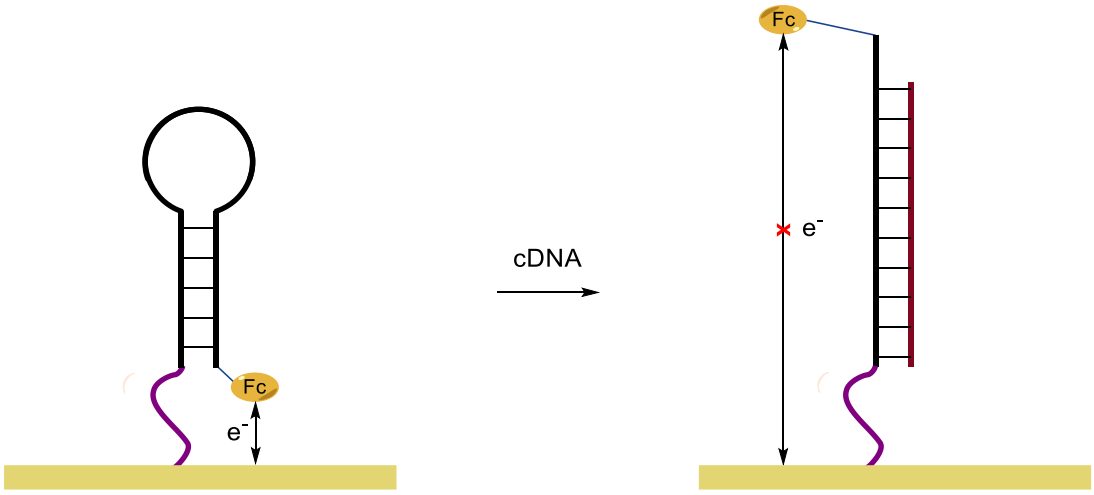
Figure 1.14 Takenaka's ferrocenyl diimide

Other DNA intercalators have been studied, exploiting the charge transfer through DNA to anthraquinonemonosulfonic acid. This allowed Gooding and Wong to discriminate between cDNA, and C-A or G-A single mismatches.[96] The distorted charge transfer between complementary and mismatched DNA resulted in a shift in reduction potential of the anthraquinone moiety.

Labelled nucleic acids have attracted more attention than label-free methods due to their increased sensitivity and lower oxidation potentials required to achieve oxidation of probes. The first electrochemical labels to be covalently attached to DNA were Os(VIII) species. In the developed conditions, osmium undergoes cyclisation reactions with pyrimidine residues in regions of dsDNA which had been distorted (*i.e.* SNPs or deletion mutations).[97,98]

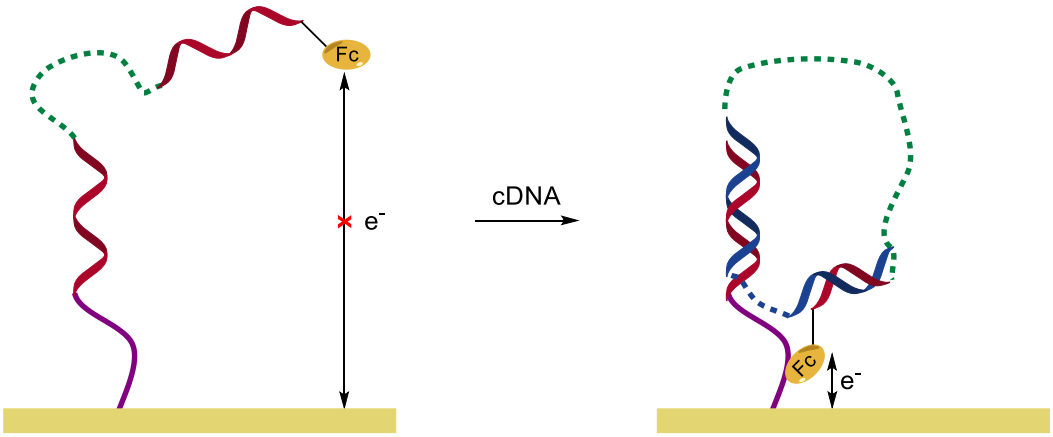
Covalent electrochemical labelling of DNA can be achieved with a range of labels including acridine,[99] anthraquinone,[51] and ruthenium based metal complexes.[86] This field, however, is dominated by derivatives of two molecules: methylene blue, and ferrocene. Methylene blue is a charged species, and as such has inherently superior aqueous solubility over ferrocene, however through conjugation of the labels to oligonucleotides, both have been used extensively in nucleic acid detection assays.

Plaxco *et al.* developed Kramer's fluorescent hairpin methodology into an amperometric electrode-based system.[100,101] The immobilised ssDNA was 5'-capped with a ferrocene label, and the probe has a self-complementary sequence, bringing the ferrocene label adjacent to the electrode in the absence of DNA, thus allowing oxidation to occur (Scheme 1.29). Upon introduction of cDNA, the self-complementary section is displaced by formation of the more stable duplex, increasing the distance between the ferrocene and the electrode, thus resulting in a signal-off current.



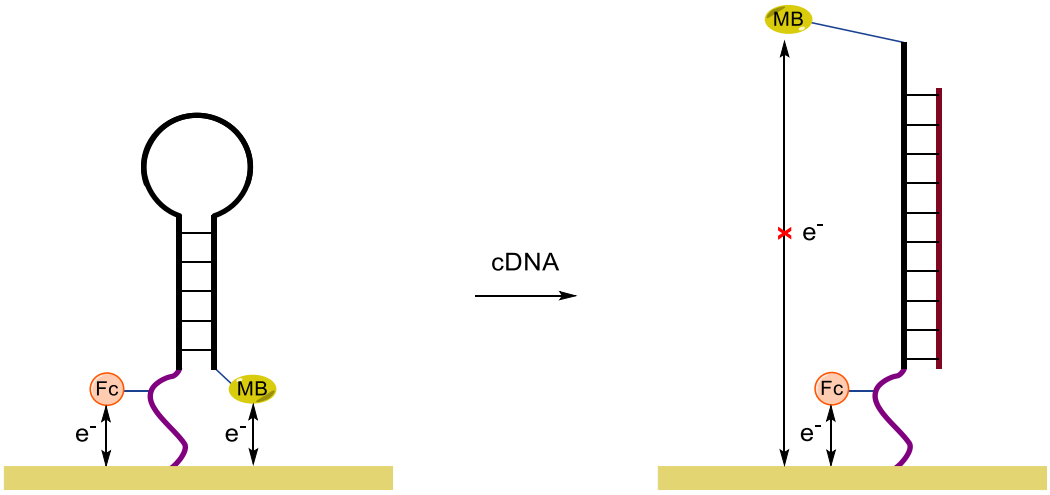
Scheme 1.29 Signal-off electrochemical molecular beacons

Grinstaff *et al.* further developed the displacement assay into a signal-on assay.[102] The introduction of a spacer unit allowed for the incorporation of electrochemically active labels, which would be off in the absence of cDNA (Scheme 1.30). Upon introduction of cDNA, the spacer has enough flexibility to allow the label to come close enough to the electrode, therefore allowing oxidation to occur.



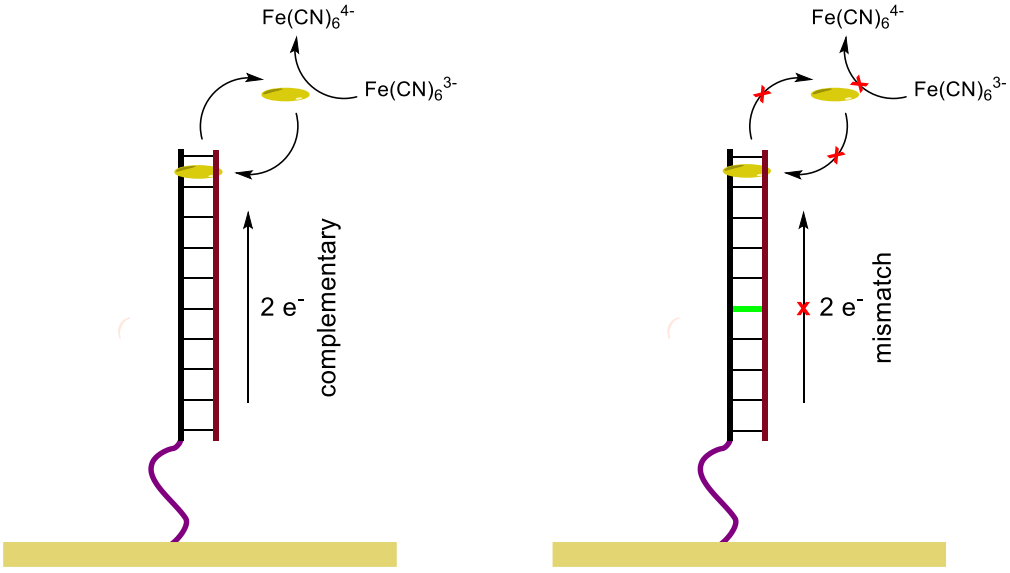
Scheme 1.30 Switch-on electrochemical detection

Sessler *et al*. further developed the displacement assays into a ratiometric detection method (Scheme 1.31).[103] By using an internal control, the signal intensity of the label could be quantified, and therefore the amount of DNA present. In this method, ferrocene was bound to the 3'- end of DNA, capped with a thiol linker, and immobilised on a gold surface. In the ssDNA state, methylene blue is in proximity to the electrode, then cDNA forces a conformational change, unfolding the probe. The ferrocene label remains the same distance from the electrode, so the integral of the MB peak can then be determined and referenced against the internal ferrocene to determine the quantity of DNA.



Scheme 1.31 Ratiometric electrochemical detection of DNA

Surface immobilised strategies can also be used to detect for single nucleotide polymorphisms (SNPs) through charge transfer from DNA to hybridised electrochemical labels. A single mismatch does not significantly affect the structure of a DNA helix, but does interfere with T_m,[104] and charge transfer due to the interruption of the extended π-system.[105–107] Electrochemical DNA intercalators can be paired with surface-immobilised probes, and in the presence of a SNP, the charge transfer switches off (Scheme 1.32).[108,109]

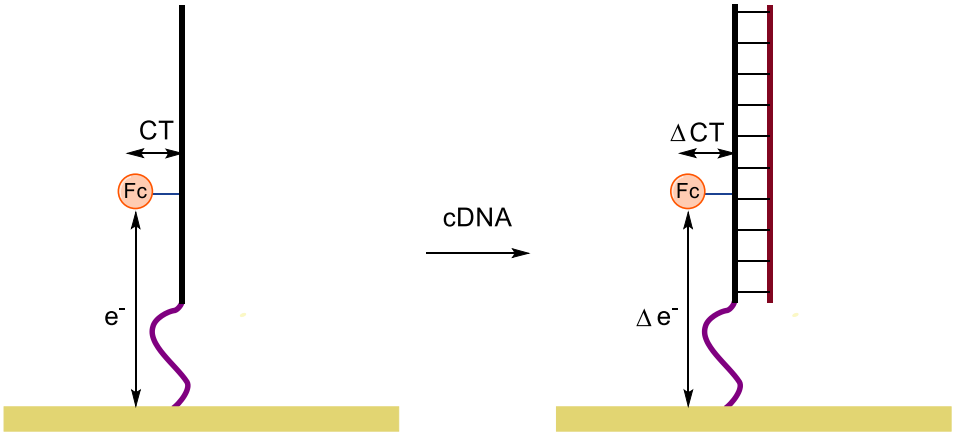


Scheme 1.32 Methylene blue/ferricyanide cycle *via* DNA charge transfer

In the presence of cDNA, the redox couple methylene blue/leucomethylene blue (LB) results in the expulsion of LB into solution as this can no longer intercalate, LB then reduces the ferricyanide in solution, allowing the reformed MB to intercalate and the complete cycle can be analysed by CV to give a fully reversible two-electron voltammogram. Upon introduction

of mismatched DNA, the charge transfer along the base stack is inefficient and as such the MB cannot be reduced, so no current is generated.

Recently, Yamana *et al*. reported the introduction of a ferrocenyl uridine into surface immobilised probes to detect mismatches for DNA.[64] The introduction of an ethene unit between ferrocene and the nucleobase allowed for the charge transfer between the nucleobase and ferrocene (Scheme 1.18, Section 1.2.5). Upon hybridisation with DNA, the level of charge transfer would change, resulting in a change in the current generated depending on whether fully complementary or mismatched DNA was present (Scheme 1.33).



Scheme 1.33 Detection of SNPs via charge transfer

1.3.3.2. Solution-based detection methods

Solution-based detection of DNA using electrochemical labels has not drawn as much attention as surface-immobilised strategies because of their comparatively reduced sensitivity, and while there are several examples in the literature of ferrocene-labelled DNA being used to study duplex formation and stability, the methodology has seldom been expanded into a full detection assay.

Surface-immobilised capture probes have advantages over solution-based detection, as the probe and label are forced into proximity to the electrode. Unmodified electrodes rely on efficient transport of the label from solution to the surface, and can be subject to fouling, therefore increasing risk of false-positives. Unmodified electrodes do however present an opportunity for use of disposable electrodes, as significantly cheaper materials can be used than those required for immobilised techniques.

Solution-based detection of DNA hybridisation has been achieved through the introduction of ferrocene units at the 5'-end and as non-nucleosidic units (Figure 1.15).[110,111] The modified oligonucleotides were studied at 25 μM, and were hybridised with complementary

sequences and detected in phosphate buffer (pH 6.8). Upon hybridisation with fully cDNA, no change in CV of the ODNs was observed, however hybridisation with complementary, but shorter sequences gave increased CV area by 45%, which was attributed to increased availability of the ferrocene at the electrode surface. The probes generated current in the range of 10^{-7} A, however this value was not baseline corrected and as such the actual sensitivity of these probes is likely to be significantly lower.

Figure 1.15 5'- and non-nucleosidic ferrocenyl DNA

Solution-based detection of ferrocenyl peptide nucleic acids (PNA) was developed by Metzler-Nolte *et al*. whereby the ferrocene was conjugated at the *N*-terminus to develop a set of probes **21**–**24** with different oxidation potentials (Figure 1.16).[112] The CV of these probes was studied, showing a shift in oxidation potential between the different PNA conjugates due to the different connectivity. The probes generated large current in the range of 10^{-6} A, however the redox behaviour of these probes was studied at 800 μM, which in a diagnostic assay would be unsuitably high.

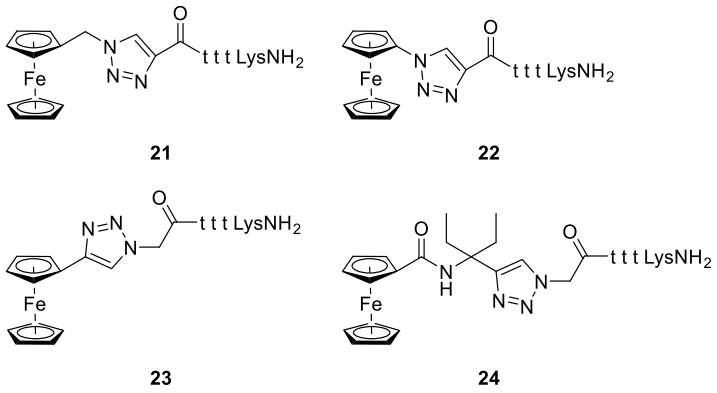
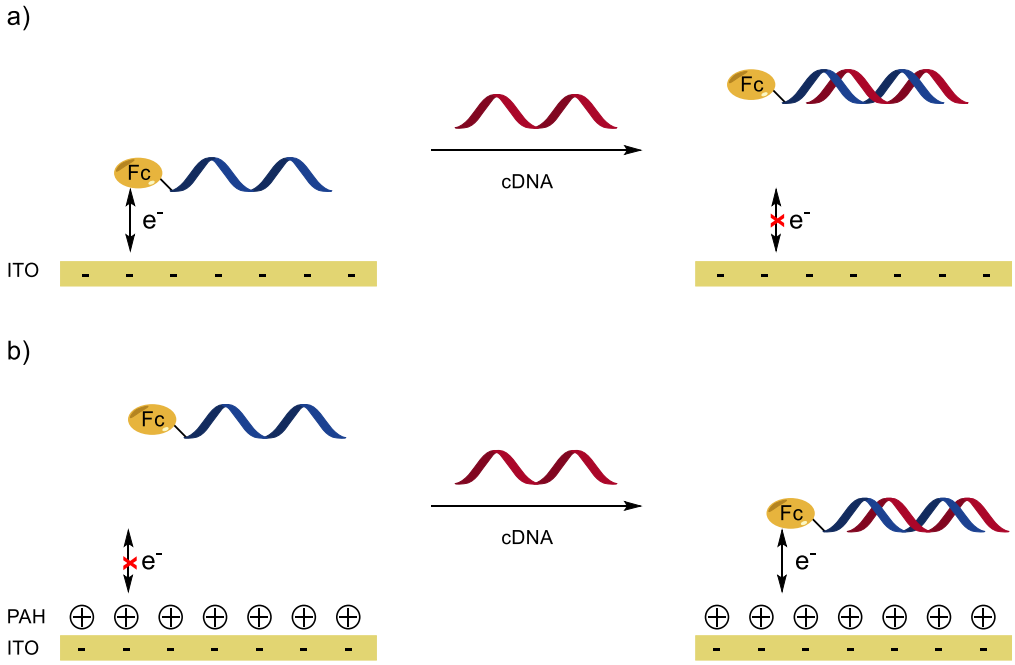


Figure 1.16 Multiplex ferrocene PNA conjugates

Hsing *et al*. developed PNA derivatives for solution-based electrochemical detection.[113] This detection strategy elegantly utilised the uncharged nature of PNA and by modification of the ITO electrode surface with positively charged poly(allylamine hydrochloride) (PAH), allowed for the design of both signal-on and signal-off strategies (Scheme 1.34).

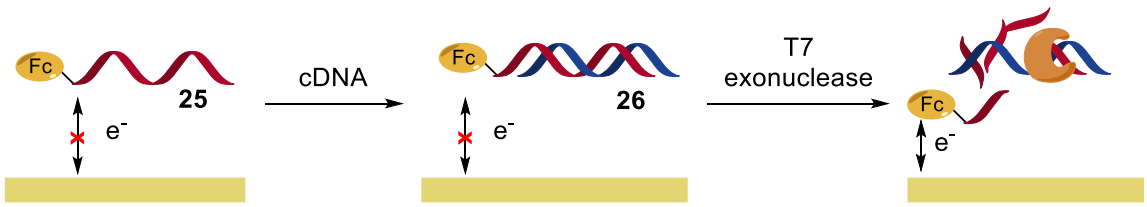
When using the unmodified surface, the neutral PNA is attracted to the negatively charged ITO electrode, allowing efficient electron transport and therefore allowing oxidation to occur at the electrode surface. Upon introduction of cDNA, the negatively charged

phosphate repels the duplex from the electrode surface, increasing the distance and therefore switching off electron transfer and any current generated. The electrode surface was then modified with PAH, forming a positively charged monolayer on the ITO surface. The single-stranded PNA probe is repelled by the electrode surface, such that it is off in the absence of cDNA. Upon introduction of the negatively charged cDNA, the duplex is attracted to the surface, allowing ferrocene oxidation to resume. The probes were studied at 1 or 0.1 μM, and hybridisation with cDNA allowed for clear differentiation at 10^{-10} A. The hybridisation event of cDNA with the probe gave a detection limit of 40 fmol.



Scheme 1.34 Immobilisation free detection of DNA

In collaboration with Atlas Genetics, previous work from this group focussed on the development of solution-based detection using 5'-ferrocenylated probe **25** (Scheme 1.35).[114–117] This was hybridised with synthetic complementary target to yield hybridised dsDNA **26**, which was then treated with T7 exonuclease, which exhibits double-strand specific activity. As such, the unhybridised probe **25** remains intact in the absence of cDNA, only producing digestion products, and therefore redox active products, when complementary DNA is present. The probes were studied at 1 μM, resulting in a signal response of 10^{-8} A upon digestion.



Scheme 1.35 Solution-based T7 exonuclease detection of DNA by ferrocene labelled probes

The limited sensitivity of solution-based detection can be overcome by coupling the reaction with efficient nucleic acid amplification techniques such as PCR, however still lacks in sensitivity compared to surface-immobilised methods. The use of surface-immobilised techniques has not found widespread use in point-of-care diagnostics largely due to the cost associated with the electrodes and the extended procedures required to immobilise probes onto the surface. Solution-based detection which relies on enzymatic digestion can be conducted with relative ease and speed compared to surface-immobilised techniques, allowing for disposable diagnostic tests to be conducted in short timeframes.

1.4. Aims

The benefits of electrochemical DNA detection over fluorescent detection devices has not yet been fully explored, and could allow for significantly cheaper diagnostic tests as well as enhanced information that can be gained from a single test. The current limitations of electrochemical detection are either the cost and labour associated (surface-immobilised methods), or the sensitivity of the devices (solution-based). The aim of this thesis was to develop a nucleic acid labelling strategy which would allow for increased sensitivity of the probes.

To achieve this, it was hypothesised that a range of ferrocene azides should be synthesised for use in the copper-catalysed azide-alkyne cycloaddition to conjugate the ferrocene label to a nucleobase, followed by incorporation of the substituted nucleoside into an oligonucleotide sequence using solid-phase synthesis (Figure 1.17). In theory, this could allow for the introduction of multiple redox active sites on a single oligonucleotide probe, and upon enzymatic digestion the probes could generate a higher electrochemical signal due to the effective increased concentration of the ferrocene label.

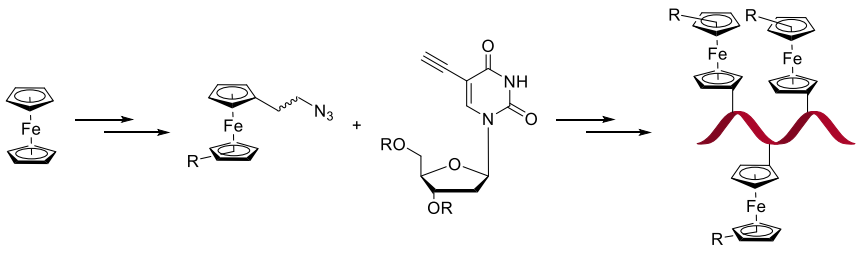


Figure 1.17 Proposed route towards ferrocene oligonucleotides

S1 nuclease was to be employed as a nuclease to digest the labelled oligonucleotides in a proof-of-concept experiment to confirm whether the probes containing multiple ferrocenes are still active towards nuclease digestion.

It was also hypothesised that further increases in electrochemical signal could be achieved *via* the use of a reductive amination strategy, to double the concentration of ferrocene without an increase in substitution of the oligonucleotide probe (Figure 1.18).

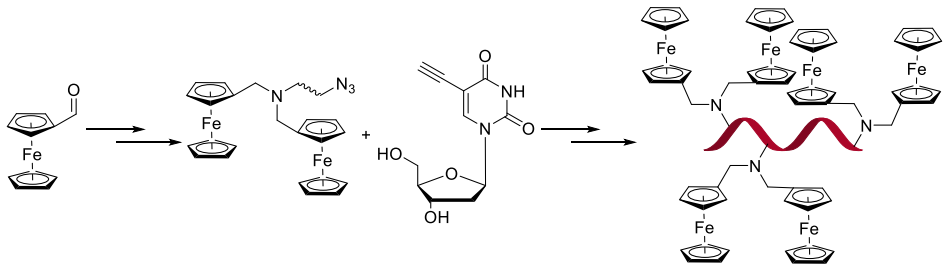


Figure 1.18 Possible increase in electrochemical signal *via* increased substitution through reductive amination reaction

These probes were also to be digested *via* S1 nuclease digestion to determine their susceptibility towards enzymatic digestion.

If the probes exhibit a suitable level of enzymatic susceptibility, the probes are to be incorporated into a dsDNA assay to study their digestion properties and determine whether the increased substitution of the oligonucleotide probes can offer improved electrochemical signals compared to traditional labelling methods using T7 exonuclease. Improvement of electrochemical signal could allow for lower quantities of probe to be used in diagnostic assays, and reduce the time required to run an assay due to potential improved sensitivity and limit of detection. Finally, the new labelling strategy could be used to explore the possibility of the use of other nucleases which could offer improved digestion properties for diagnostic devices.

2. Design, synthesis, and testing of mono-ferrocenylated nucleosides

2.1. Introduction

Ferrocene was first reported by Kealy and Pauson in 1951,[118] and its structure was elucidated a year later by Wilkinson and co-workers.[119] The organometallic compound is air and moisture stable and undergoes a one-electron oxidation, and the ferrocene core can tolerate a range of organic chemistry reaction conditions. It is relatively stable to both acidic and basic conditions, and the cyclopentadienyl (Cp) rings can be manipulated through S_NAr and S_EAr chemistry, as well as undergoing selective deprotonations to synthesise a host of highly functionalised cores. Ferrocene cores can be synthesised from iron (II) chloride with functionality pre-installed through the use of functionalised Cp rings. These properties have enabled ferrocene to be applied in many areas of chemistry, including responsive materials,[120] asymmetric ligands,[121] electrochemical references,[122] petrochemicals,[123] and pharmaceuticals,[124] and extensively as a redox label in DNA detection.[125]

Electrochemical detection is typically conducted *via* voltammetric analysis. One of the fundamental principles that allows ferrocene to be used in such applications is its controlled, yet variable, oxidation potential. By manipulating the oxidation potential of the iron centre and forcing a change upon recognition of an analyte, a reaction can be monitored. It is possible to develop switches where the probe has a discrete E_{ox} in the absence or presence of the analyte.[126] This is known as ratiometric analysis, whereby the integral of the two peaks are compared and therefore quantify the amount of probe present (Figure 2.1). This is a form of duplex detection, which is the simultaneous detection of two signalling moieties.

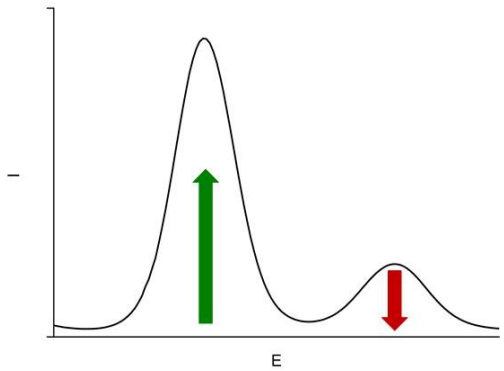
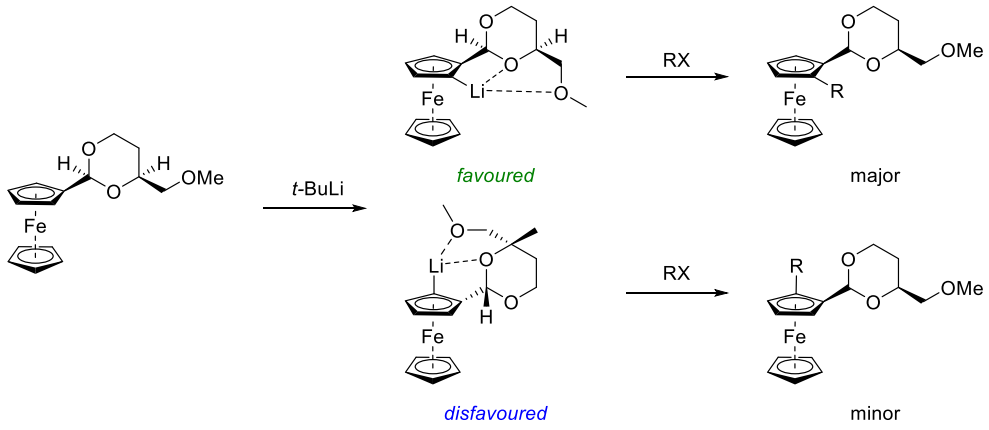


Figure 2.1 Ratiometric analysis of two distinct electrochemical species

Substitution of the Cp rings enables modification of the oxidation potential, and therefore allowing for multiplexing capabilities for electrochemical applications. The introduction of electron-donating groups on the Cp rings puts more electron density onto the iron core, therefore making it easier to lose an electron, making the oxidation potential more negative. The opposite is also true with electron withdrawing groups, where the electron density is moved away from the Cp rings, therefore less density on the iron centre, leading to a higher oxidation potential.

Deprotonation of ferrocene and quenching with an electrophile is one of the most simple and general ways in which to create substituted ferrocenes. A variety of functionality can be installed, allowing for a range of reactions to be performed. For example, quenching lithioferrocene with DMF yields ferrocenecarboxaldehyde. The aldehyde can then be subjected to standard carbonyl chemistry. This strategy allows for further derivatisation of ferrocene along a single carbon chain.

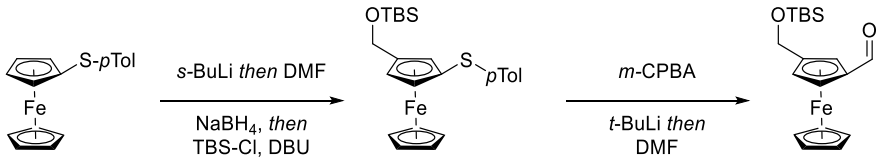
It is desirable to introduce multiple substituents on the ferrocene core, which are typically substituted in 1,2- or 1,1'- patterns. 1,2-substitution generates chiral ferrocenes *via* planar chirality. The first chiral ferrocenes synthesised were resolved *via* the formation of diastereoisomeric salts,[127] however later methods developed allowed for asymmetric lithiation,[128,129] and therefore direct access to enantiomerically enriched samples. These methods of lithiation rely on the formation of more stable conformers, whereby one orientation of the auxiliary will be disfavoured due to steric clash. Kagan *et al*. further developed the strategy to generate removable, chiral auxiliaries therefore allowing further modification after introduction of stereochemical information (Scheme 2.1).[130]



Scheme 2.1 Origin of stereochemical control in Kagan's chiral ferrocene

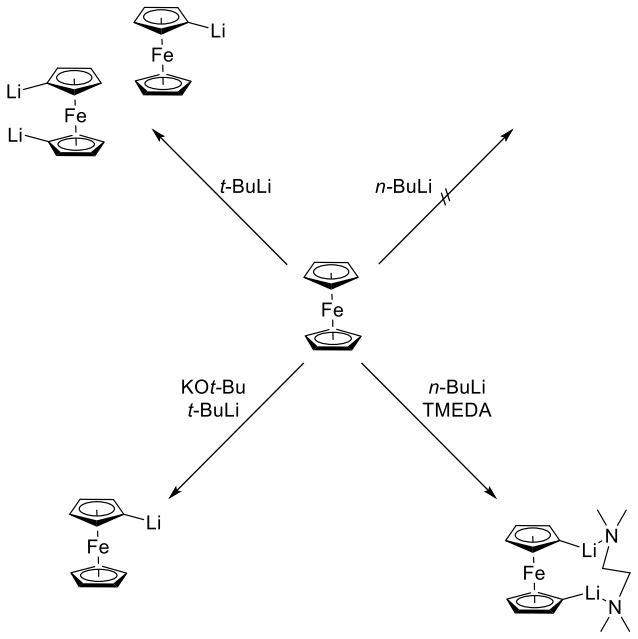
1,3-substitution patterns have been synthesised through the diacetylation of ferrocene,[131] however this was not developed as a regioselective reaction and the products require separation from their 1,1'-regioisomers. The use of sulfide as a directing group allows for

meta-selective deprotonation,[132] and subsequent quenching with electrophile, generates much improved yields. Oxidation of the directing group to the corresponding sulfoxide, followed by lithium-sulfoxide exchange allowed for access to 1,3-substituted ferrocenes in much improved selectivities and yields (Scheme 2.2). This strategy has, however, received significantly less attention over the years than the 1,2-derivatives and as such are scant literature references in comparison to 1,2-counterparts.



Scheme 2.2 1,3-substitution of ferrocene

Di-lithiation allows access to regiospecific 1,1'-substituted ferrocenes. This intermediate can be quenched with a range of electrophiles. Extensive studies by Sanders and Mueller-Westerhoff developed general strategies towards mono- and di-lithiated species (Scheme 2.3).[133] To access the mono-lithiated species, a mixture of *t*-BuLi and KO*t*Bu are employed, whereas di-lithiation occurs through treatment of ferrocene with *n*-BuLi and TMEDA. 1,1'-bromination provides access to asymmetrically substituted Cp rings through subsequent stoichiometry-controlled lithium-halogen exchange.



Scheme 2.3 Mono- and di-lithiations of ferrocene

Modification of the electronics of the Cp rings allows for multiplex capability to allow for detection of multiple targets in a single sample. Major modifications are achieved through

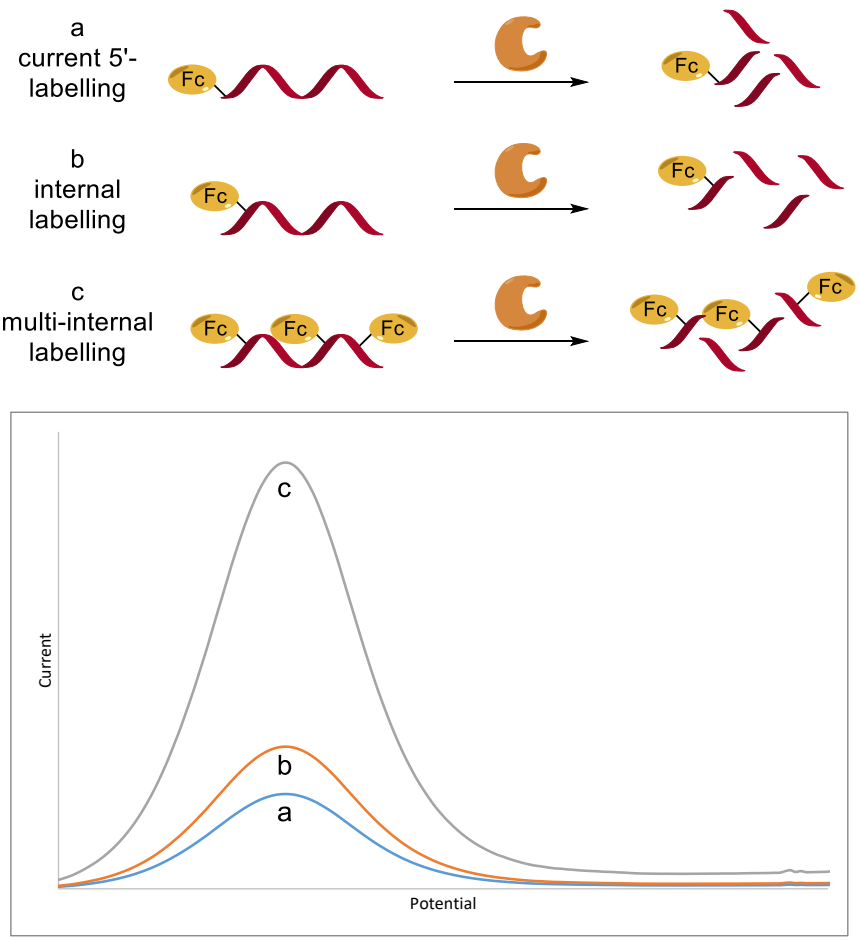
direct substitution of the Cp ring. For example, increasing the electron density through replacement of one Cp for pentamethylcyclopentadiene (Cp*) results in a large shift compared to the unsubstituted version. Electron poor derivatives can be created through substitution of the ring with groups such as halogens or sulfoxides.

Minor modification of the oxidation potential can be achieved through varying the chain length attached to the ferrocene.[134] Increased hydrophobicity of the ferrocene can increase the oxidation potential by up to 100 mV, allowing for fine-tuning of the ferrocene after the rough oxidation potential has been achieved through substituent incorporation.

The majority of electrochemical DNA detection is achieved through surface modification of gold electrodes. Whilst offering higher sensitivity than their screen-printed counterparts, the increased cost of gold electrodes poses a significant stumbling block for widespread use of electrochemical DNA diagnostics. Electrochemical detection has huge benefits over fluorescent detection, however they cannot live up to their full potential if the materials continue to be prohibitively expensive.

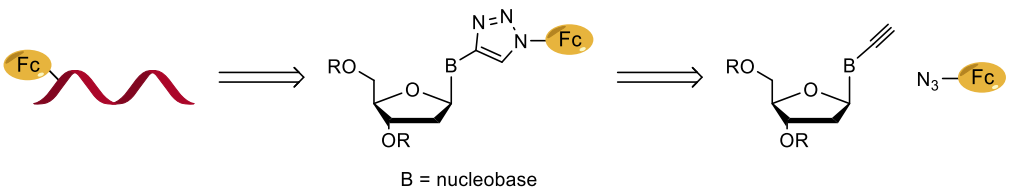
As such, solution-based, screen-printed electrodes offer a cheaper alternative and could allow for widespread use of such electrochemical DNA devices. However, solution-based DNA detection often suffers from low sensitivity of probes, although there are several examples in the literature where, when combined with amplification techniques, solution-based detection is a suitable alternative to surface-immobilised versions.

It was hypothesised that through the introduction of multiple redox active units along an oligonucleotide backbone, the signal could be increased significantly and improve the properties of such solution-based devices.



Scheme 2.4 Proposed increase in signal from multi-internal labelling

Upon conjugation of multiple units of the same redox label, the signal could be amplified due to increased concentration of the signalling molecule. The CuAAC reaction has been used extensively in DNA modification, and allows for late-stage modification, which could allow for general conditions to synthesise a range of redox active nucleosides. Through late-stage functionalisation, the required connectivity of the nucleosides can be installed (*e.g.* activating and protecting groups), before the desired ferrocene is installed. This allows for the incorporation of ferrocenes with varying redox potentials without significant deviation from a general synthesis.



Scheme 2.5 Retrosynthetic analysis of target nucleic acid

Given the potential application of these probes is for diagnostic assays, the probes are required to be synthesised with high purity. As such, the phosphoramidite method, rather

than enzymatic incorporation of dNTPs was chosen. It was also deemed necessary to incorporate the redox active label prior to solid-phase synthesis, to ensure complete conjugation of the ferrocene to the oligonucleotide, and remove any requirements for additional purification.

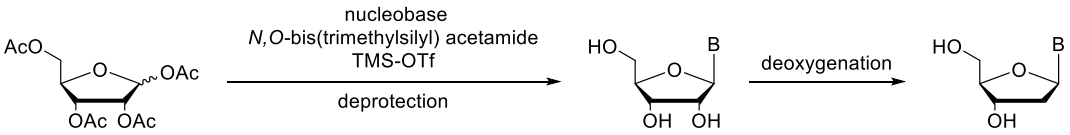
As such it was decided to investigate the synthesis of a library of ferrocenyl azides, and conjugate these directly to alkyne-modified 2'-deoxynucleosides. This would allow for the incorporation of these labels in a sequence-specific manner, allowing evaluation of possible strategy to improve electrochemical signal in a diagnostic DNA assay.

2.2. Nucleobase modification

Due to the easier chemical modification of purine nucleobases, 2'-deoxyuridine was chosen as the initial nucleoside to modify, as the Watson-Crick pairing is the same as thymidine, but possesses a vacant 5-position in which to introduce functionality. To this extent, initial investigations focussed on the introduction of an alkyne moiety in the 5-position of the purine bases.

2.2.1. Uridine substitution

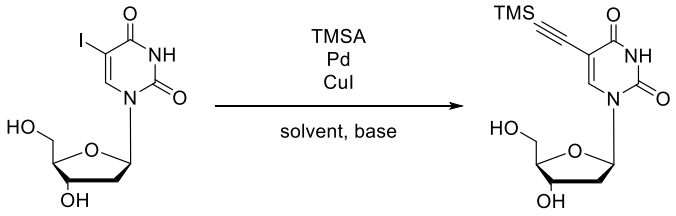
It was initially investigated to construct the substituted uridine from uracil, installing both the alkyne and *N*-glycosidic bond. The initial reactions were conducted, showing successful incorporation of both the alkyne moiety and successful installation of the glycosidic bond. To ensure the correct β-anomer is synthesised during the formation of the glycoside, ribose must still possess the 2'-OH be used to assist with neighbouring group participation. As such, after the formation of the desired nucleoside, the 2'-hydroxyl residue must be removed to enable use of the monomer in DNA (Scheme 2.6).



Scheme 2.6 Synthesis of 2'-deoxynucelosides

To achieve this, the radical Barton-McCombie deoxygenation reaction was to be attempted to form the desired 2'-deoxuridine core.[135] However, this introduces additional steps in a reaction that is not typically high yielding, and as such, it was sought to use the pre-formed nucleosides as starting material.

5-iodo-2'-deoxyuridine, although expensive as a starting material (£5/g in a moderate length linear synthesis, *cf.* uracil at £0.40/g), was purchased and used as starting material for the introduction of the required functionality. The cross coupling between 5-iodo-2'-deoxyuridine and trimethylsilyl acetylene has been widely reported in the literature.[136,137]

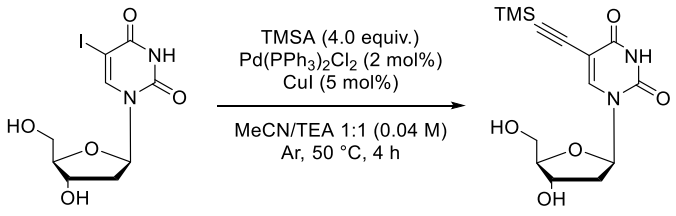


12 25

Entry	TMSA (equiv.)	Pd source (mol%)	CuI (mol%)	Base (equiv.)	Solvent (conc.)	Temp (°C)	Time (h)	Atm.	Yield (%)
1[138]	5.0	$Pd(PPh_3)_2Cl_2$ (2.3)	8.8	-	DMF/Et_3N 2:1 (0.13)	rt	6	Ar	38
2[139]	2.0	$Pd(PPh_3)_2Cl_2$ (2.0)	7.8	-	Et_3N (0.25)	50	4	N_2	24
3[140]	4.0	$Pd(PPh_3)_4$ (10)	10	Et_3N (2.0)	DMF (0.2)	55	17	Ar	53
4[141]	4.0	$Pd(PPh_3)_2Cl_2$ (2)	5	-	MeCN/Et_3N (0.04)	50	3.5	Ar	71
5[142]	1.5	$Pd(PPh_3)_4$ (5.5)	9.4	Et_3N (1.5)	DMF (0.13)	50	2	N_2	28

Table 2.1 Cross-coupling screening reactions

After attempting several procedures (Table 2.1) it was found the procedure reported by Cristofoli *et al.* gave the desired compound in the highest yield.[141] It was found that thorough deoxygenation of the system was required to enable high yielding reactions, with moisture sensitivity not a concern (Scheme 2.7). The reaction was scaled up to 20 mmol, with no significant reduction in yield (73%).

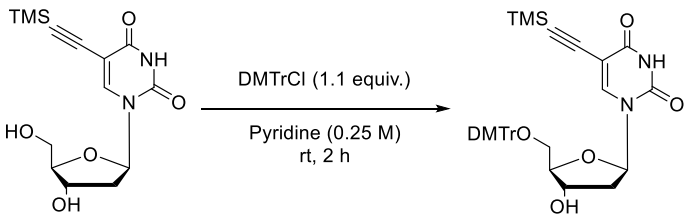


12 **25**, 86%

Scheme 2.7 Sonogashira cross-coupling of nucleoside

The nucleobase must be activated for solid phase synthesis, and nucleophilic sites that would result in unwanted polymeric products need to be protected. DMTr phosphoramidites are standard coupling reagents used in solid phase synthesis.[32] As such, the 5'-hydroxyl group must be protected in such a way that allows for deprotection under mild conditions. The tritylation of nucleosides is ubiquitous in the literature and as such, a

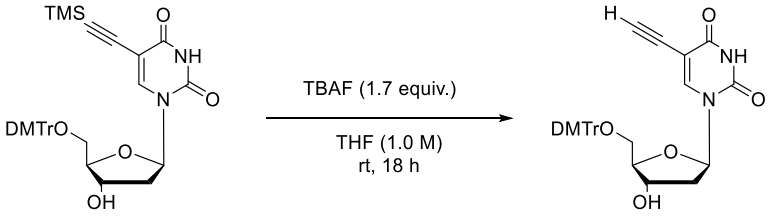
range of general procedures were available from which to choose.[143] The key findings from this were that to maximise yields, the uridine should be azeotroped with anhydrous pyridine prior to use, and ensure that only freshly distilled pyridine is used as solvent.



25 **26**, 60%

Scheme 2.8 5'-*O* protection of 2'-deoxyuridine

Removal of the TMS group to afford the desired 5-ethynyl-5'-*O*-DMTr-2'-deoxyuridine was attempted *via* potassium carbonate in methanol.[144] This yielded the desired final compound in a 39% yield, however it was found that the use of TBAF in THF proved more reliable with higher yields and as such was the chosen method (Scheme 2.9).[145]



26 **27**, 58%

Scheme 2.9 Alkyne deprotection

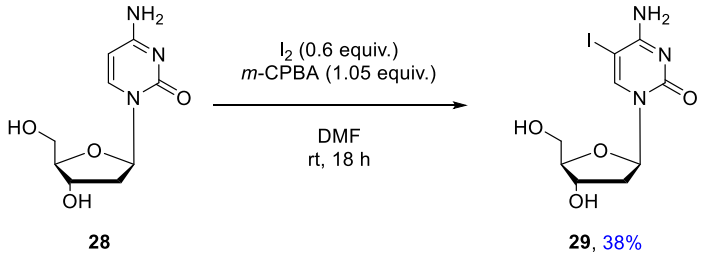
Upon successful synthesis of **27**, one of the key intermediates for the CuAAC had been generated. This intermediate allows for the inclusion of any azide that has moderate acid and base stability. Attention was turned to other nucleosides to increase the labelling versatility of the ferrocenyl nucleosides.

2.2.2. Cytidine substitution

2'-deoxycytidine was investigated as a new coupling partner. In contrast to 5-iodo-2'-deoxyuridine, 5-halo-2'-deoxycytidines were prohibitively expensive for use as a starting material, either as the iodide or bromide, and as such the direct iodination of cytidine was investigated.

The use of elemental iodine and *m*-CPBA afforded the desired compound in 38% yield after column chromatography.[146] Due to the high polarity of the product, very polar solvent mixtures were required, which included the addition of ammonia and water the mobile

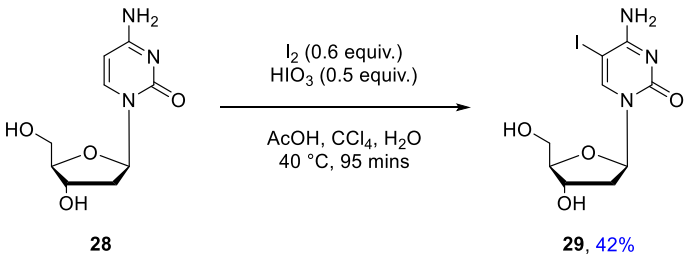
phase. Due to the high proportion of immiscible aqueous and organic solvents, this resulted in difficult separation from by-products. Recrystallisation conditions were investigated however no suitable conditions were found. The reaction was also difficult to reproduce, as following the reaction *via* TLC showed no conversion to desired product when attempting the reaction on other occasions.



28 **29**, 38%

Scheme 2.10 Iodination of 2'-deoxycytidine with *m*-CPBA as oxidant

In the presence of a mixture of iodine, iodic acid, carbon tetrachloride, acetic acid and water, 2'-deoxycytidine was able to cleanly undergo iodination in the 5-position (Scheme 2.11).[147] The reaction mixture was quenched with water, and co-evaporated with methanol until the reaction was colourless. This removed much of the iodine from the mixture, allowing the desired product to be successfully recrystallised from water, adjusted to pH 10 with 1M NaOH.



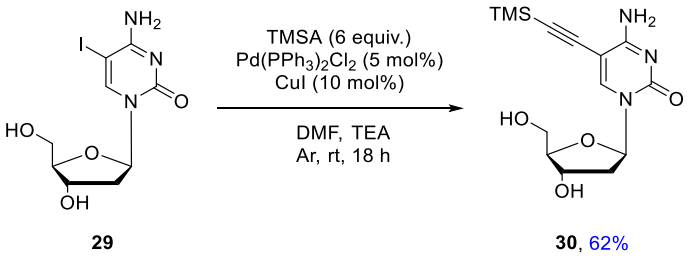
28 **29**, 42%

Scheme 2.11 Iodination of 2'-deoxycytidine with HIO_3 as oxidant

Although neither yield from the reaction was particularly high, the iodic acid reaction resulting product was of a higher purity by ^{1}H NMR spectroscopy and as such was chosen as a more suitable method for the preparation of this starting material.

The installation of an alkynyl group was then attempted. As was found with the uridine derivative, there are many examples in the literature to install the ethynyl moiety. After attempting several methods, however, it was found that the isolation of **30** was particularly difficult due to its highly polar nature. Crude ^{1}H NMR spectroscopy showed success for several sets of conditions, however it was difficult to isolate **30** in high purity. Finally, a set

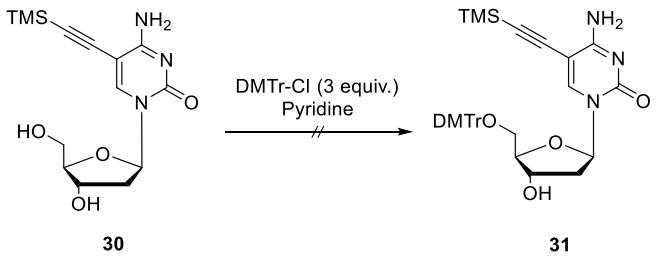
of reaction and purification conditions were found which provided **30** in high purity and reasonable yield (Scheme 2.12).[148]



Scheme 2.12 Sonogashira cross-coupling of 2'-deoxycytidine

After the successful synthesis of the alkynylated 2'-deoxycytidine, the protection of the exocyclic amine and 5'- hydroxyl was attempted. The amine must be protected due to the phosphoramidite coupling method, as the exocyclic amine is nucleophilic. The amine would therefore couple with the incoming phosphoramidite. This would create a base that would be conjugated *via* the nucleobase rather than the ribose, hence causing problems with Watson-Crick binding.

5'-*O*-DMTr protection was attempted under the same conditions as the uridine derivative. Crude TLC and ^{1}H NMR spectroscopy indicated conversion to the desired product, and after work-up and purification the desired spot was isolated, however the product was not stable, showing decomposition to a number of spots by TLC and growing of a large number of peaks present in ^{1}H NMR analysis (Scheme 2.13).



Scheme 2.13 Attempted protection of 2'-deoxycytidine

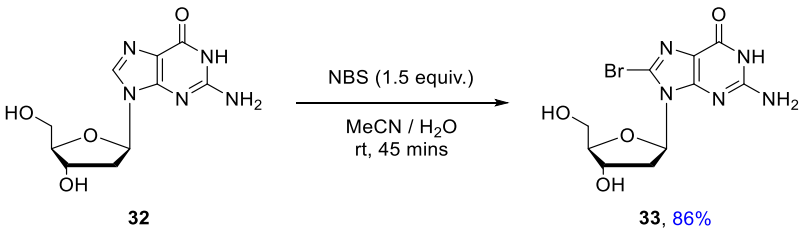
It was therefore attempted to telescope the reaction, and protect both the hydroxyl and amine groups before purification by first installing the DMTr group, followed by acyl protection of the exocyclic amine (Scheme 2.13). Crude TLC indicated the 5'-OH reaction proceeded with full conversion. The reaction was quenched with MeOH to remove any remaining DMTrCl, and the crude mixture was subjected to protection with either acetyl or DMF protecting groups. Unfortunately, although the first protection of the alcohol appeared to proceed without issue, the introduction of a second protecting group was not fruitful. On

addition of either acetyl chloride, acetic anhydride, or DMF-DMA, crude TLC indicated a large number of new spots, many of which were fluorescent under long-wavelength UV light. Purification *via* silica gel chromatography was attempted, however highly polar mixtures of apolar solvents and MeOH did not give suitable separation. Water and/or ammonia was added to the eluents to improve separation, and although this helped to isolate single spots by TLC, upon analysis by NMR it was found that once again, the product was not clean. As such, investigations towards the synthesis of an alkynyl substituted cytidine were terminated.

2.2.3. Purine bases

After the unsuccessful synthesis of the alkynyl cytidine derivative, it was attempted to functionalise both 2'-deoxyadenosine (dA) and 2'-deoxyguanosine (dG). As was the case with halogenated cytidine, 8-halogenated dG and dA were prohibitively expensive. As a result, the direct halogenation of both was required and a number of literature methods were attempted.

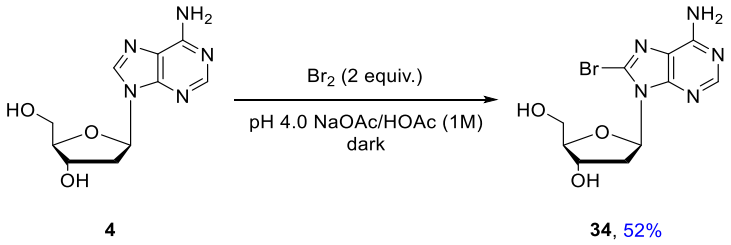
The bromination of guanosine was first attempted, and the first conditions attempted proved successful. The was treated with *N*-bromosuccinimide in a mixture of acetonitrile and water, evaporated to dryness and recrystallised from acetone at –20 °C for 48 hours.[149] This provided the desired product in a good yield, with no further purification required.



Scheme 2.14 Bromination of 2'-deoxyguanosine

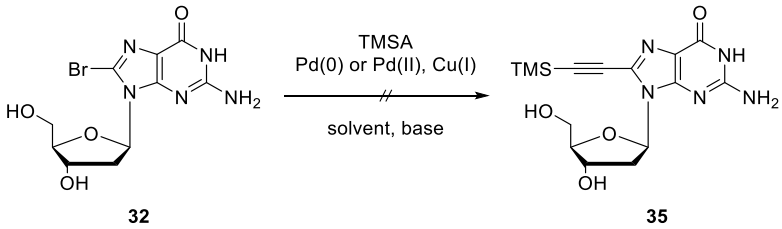
The bromination of adenosine was also attempted, however this required a more thorough investigation. After the success of guanosine with NBS, this was repeated on the adenosine substrate, however no conversion to desired product was observed. Elemental bromine was then employed on the substrate, as a literature search showed that many methods utilise aqueous bromine in acidic sodium acetate buffer. Of the many procedures reported, most vary slightly in concentration of buffer used or in work-up procedure. After attempting several literature conditions, it was found that bromine in pH 4.0 sodium acetate/acetic acid 1.0 M buffer was suitable, however it was important to run the reaction at a low concentration of 2'-deoxyadenosine of 0.05 M and ensure that the reaction was run in the

absence of light. The desired product was isolated *via* co-extraction of acetic acid and the product with CH_2Cl_2.



Scheme 2.15 Bromination of 2'-deoxyadenosine

After the successful isolation of the pyrimidine bases, the Sonogashira cross-couplings were attempted. Literature procedures for the cross-coupling with Pd(II) species were followed, however crude TLCs showed no conversion, or very weakly fluorescent new spots forming. In cases where the reaction did show new spots by TLC, it was attempted to isolate the new product *via* silica gel chromatography.



Scheme 2.16 Attempted cross-coupling of 2'-deoxyguanosine

In a similar manner to 2'-deoxycytidine, the pyrimidine nucleosides were incredibly polar and it was difficult to isolate new spots. After some attempts, conditions were found which resulted in crude ^{1}H NMR and MS indicating desired product, however upon purification, only trace amounts of low purity were obtained. Upon introduction of Pd(0) species, the reactions resulted in the formation of highly fluorescent species for both dG and dA. Due to the strongly coordinating nature of the purines and their increased number of coordinating atoms compared to the pyrimidines, the adenosine and guanosine is likely getting too strongly adsorbed to the silica and can no longer be isolated because of the strong H-bonding array that is formed with the solid phase.

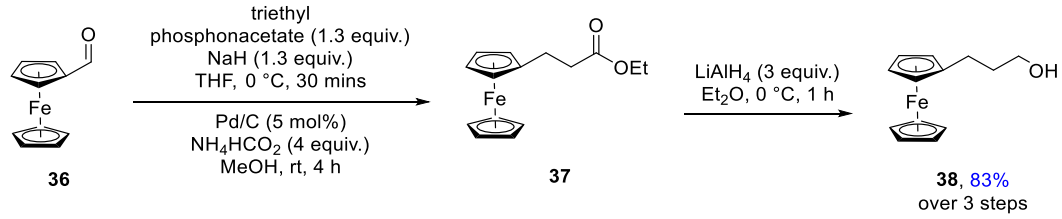
Due to the difficulties encountered during protection of the cytidine derivative, it was decided that while it was desirable to include additional nucleobases in the study, in the interest of time the route towards redox active pyrimidines would be abandoned.

Disappointingly, this resulted in the isolation of deoxyuridine as the only modified nucleoside. This somewhat limits the biological targets that can be selected to those that are

rich in adenosine (due to the A-T binding), however it was still possible to conduct a thorough investigation into the effects of internal labelling on enzymatic digestion.

2.3. Ferrocenyl azide library development

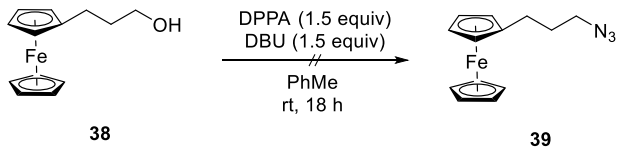
With access to the patents and intellectual property of Atlas Genetics, their library of ferrocene compounds was studied to find a suitable ferrocene label to conjugate to the uridine. The key decisions to make were regarding the length of the linker to tether the ferrocene, and the desired oxidation potential of the substrate. As the following studies were a proof-of-principle investigation, it was decided that an unsubstituted ferrocene would be used with no additional functionality on the cyclopentadiene rings. The length of the linker to be used was also of importance, as the additional aromatic triazole formed in any click reaction between the base and the ferrocene would add additional length to the linker. It was hypothesised that too short a linker may disrupt base pairing of the nucleobases, however too long a tether made allow the ferrocene to wrap back around and interfere with the binding of DNA to double strands. As such, it was decided to use a simple ferrocene linker, 3-ferrocenylpropan-1-ol, **38**.



Scheme 2.17 Synthesis of 3-ferrocenylpropan-1-ol

To achieve this, ferrocenecarboxaldehyde was first subjected to the Horner-Wadsworth-Emmons reaction. The resultant enone was isolated quantitatively after a quick work up and trituration with hexane. The enone of the resulting product was hydrogenated, again resulting in quantitative conversion and isolation of clean ferrocenyl ester. Finally, the ester was reduced and purified to prepare the desired compound in an 83% yield (Scheme 2.17).

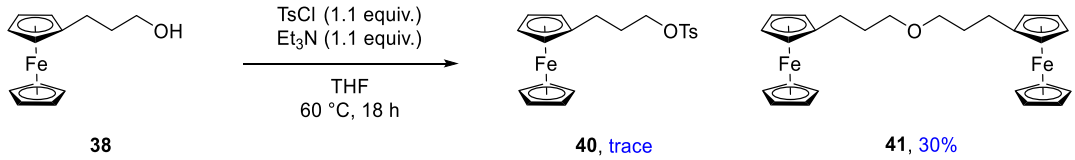
It was then necessary to convert the alcohol into the corresponding azide. This was first attempted *via* the use of diphenylphosphoryl azide (DPPA). This did not, however, lead to desired product. The DPPA did indeed undergo electrophilic attack from the nucleophilic oxygen, however the reaction did not proceed past the intermediate phosphinate, as the resulting azide nucleophile did not displace the phosphinate to give desired product. It was thought that additional azide source in the form of sodium azide could help the reaction to proceed, however even three equivalents of sodium azide did not help form the desired product.



38 39

Scheme 2.18 Attempted direct azidation of ferrocenyl alcohol

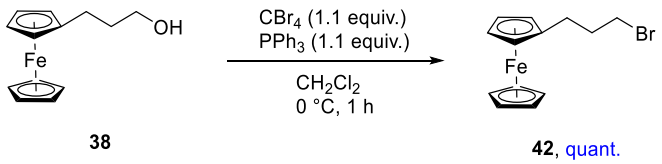
A two-step reaction was instead investigated, whereby the alcohol would first be converted into a suitable leaving group and then undergo nucleophilic substitution. The first leaving group that was utilised was a tosylate. The alcohol was treated with tosyl chloride and triethylamine, however leaving the reaction overnight at 40 °C resulted in no new compounds being formed *via* TLC analysis. The temperature was further increased to 60 °C, after which point new spots began to appear *via* TLC. A work-up was carried out and the new spot was isolated, however it was found that while the tosylate had indeed been formed *in situ*, the resulting product was more reactive than the tosyl chloride, and as such the remaining alcohol attacked the tosylate intermediate **40** to yield the dimer **41**. The reaction was repeated in higher dilution and increased equivalence of tosyl chloride, however the dimer was still formed as the major product.



38 **40**, trace **41**, 30%

Scheme 2.19 Tosylation of ferrocenyl alcohol

The installation of a bromide leaving group was therefore investigated. This was achieved through the Appel reaction, treating the alcohol with carbon tetrabromide and triphenylphosphine. No evidence of the ferrocenyl dimer **41** was observed by TLC or ^{1}H NMR analysis.



38 **42**, quant.

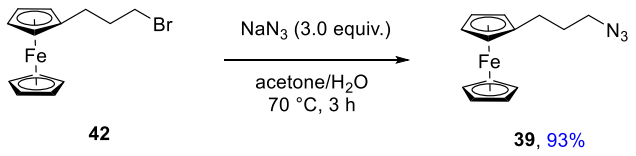
Scheme 2.20 Bromination of ferrocenyl alcohol

The initial reaction gave the desired product in a 73% yield, however conducting the reaction in anhydrous solvent and cooling for the addition of PPh_3 increased the yield. The work-up procedure was further improved after trituration of the reaction mixture with hexanes to precipitate out the phosphine oxide formed in the reaction, enabling the desired

compound to be isolated quantitatively. This compound was found to be light sensitive - if the vial was stored on the bench for two days, a powdery, brown residue began to form around the glass. The decomposing mixture could, however, be passed through a silica plug to clean the remaining material and then be used in further reaction without issues.

A suitable method to convert the bromide to the azide was then investigated. Sodium azide was employed as the azide source. Initial conditions used sodium azide in a CH_2Cl_2:H_2O mixture with TBAI as a phase transfer and nucleophilic catalyst, however the product was only isolated in 40% yield, and due to safety concerns during potential scale up of mixing chlorinated solvents and azide species, other methods were investigated. Literature searches of the conversion primary bromides into azides gave new conditions of using an excess of sodium azide in ethanol under reflux, which again only gave the product in 50% yield. Considering the reaction is a simple S_N2, it would be expected to undergo conversion to product in higher yields and as such further methods were investigated.

The third method that was attempted also used an excess of sodium azide in a 1:1 mixture of acetone/water (Scheme 2.21). Stirring at room temperature overnight gave no conversion, however heating the mixture to 70 °C dramatically increased the rate of the reaction, and after heating for 2 hours no starting material remained. After a simple work-up and filtration through silica, the desired ferrocenyl azide **39** was obtained in 95% yield.



42 **39**, 93%

Scheme 2.21 Conversion of ferrocenyl bromide to azide

2.3.1. Substituted ferrocenes

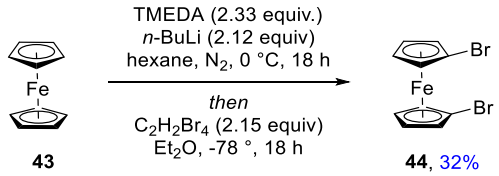
Through changing the electron density on the ferrocene ring, the oxidation potential can be finely tuned. By introducing electron-donating or electron-withdrawing groups, the number of signals that can be incorporated in the range of the electrodes is increased due to the small peak width of the ferrocene based system. This allows for the simultaneous detection of different DNA targets *via* the conjugation of labels with different oxidation potentials to different probe sequences.

The research carried out by Atlas Genetics has previously identified a range of substitution patterns with varying oxidation potentials that are suitable for multiplexing. As such, the goal in this chapter was not to identify a multiplex capability, but to investigate whether incorporating “mass-spectrometry labels” would be applicable. As such, electrophiles with isotopic patterns visible in mass spectrometry were chosen to allow for the enzymatic

digestion products of probes to be studied *via* mass spectrometry, with the goal of making the digested oligonucleotides easier to identify due to the isotopic pattern present. This could, in theory, provide insight to action of various chosen enzymes, and therefore inform the optimal labelling sequence of probes.

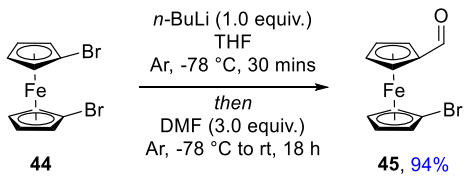
To modulate the oxidation potential of ferrocene, the functionality on the cyclopentadiene rings must be modified. The use of 1,1'-dibromoferrocene as starting material allows for the introduction of two different functional groups through lithium-halogen exchange, yielding 1,1'-disubstituted ferrocenes. The aldehyde handle on a ferrocene is particularly useful, as this allows a range of chemistry to be carried out immediately after functionalisation, as utilised with the formation of the first ferrocenylazide label through the Horner-Wadsworth-Emmons reaction. It can also be reduced to the alcohol and the benzylic alcohol can then be etherified though the Lewis-acid catalysed substitution with diols.

To this extent, 1,1'-dibromoferrocene was synthesised according to the literature procedure.[133] Additionally, the reaction is easier to conduct on a large scale (100 mmol) allowing stocks of starting material to be synthesised (Scheme 2.22). Despite having largely different electronic properties, halogenated ferrocenes are difficult to separate by chromatography. An elegant method of purification developed by Long and co-workers exploited the differing oxidation potentials by using iron (III) chloride solution to oxidise any remaining unreacted and mono-brominated ferrocene.[150] The desired product was then recrystallised from hot methanol to remove any unreacted tetrabromoethane.



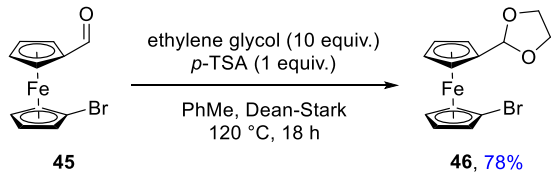
Scheme 2.22 Synthesis of 1,1'-dibromoferrocene

1,1'-dibromoferrocene was then converted into 1'-bromoferrocenecarboxaldehyde through treatment of the dibromide with *n*-BuLi, followed by quenching with DMF to yield the desired aldehyde.



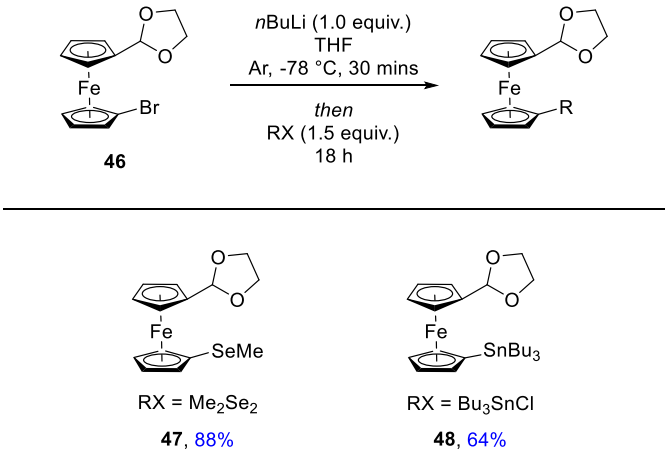
Scheme 2.23 Lithium-halogen exchange and quenching of dibromoferrocene

This aldehyde itself has enough functionality to affect the oxidation potential of ferrocene, however it is desirable to introduce a range of functionality to allow for further multiplexing capability. As such, the aldehyde was protected as the 1,3-dioxolane to allow for further lithium-halogen exchanges to occur.



Scheme 2.24 Acetal formation of ferrocenecarboxaldehyde

With the protection of the electrophilic site complete additional substitution could now be performed. To synthesise the substituted ferrocenes, the standard electrophilic substitution was carried out *via* lithium-halogen exchange followed by quenching with the desired electrophile. The reactions must be carried out in a thoroughly deoxygenated Schlenk flask, otherwise the reaction often failed or gave very low yields, resulting in either the protonated species, or unidentified decomposition products. The reaction is simple to track, as after deprotonation with *n*-BuLi, the FcLi precipitate forms as it is insoluble in THF at low temperatures. This is indicative of a successful lithium-halogen exchange and as such can be used as an early indication of the success of the reaction. Work up with sodium hydroxide and purification *via* silica gel chromatography then generated the newly substituted ferrocenes in good, reproducible yields.



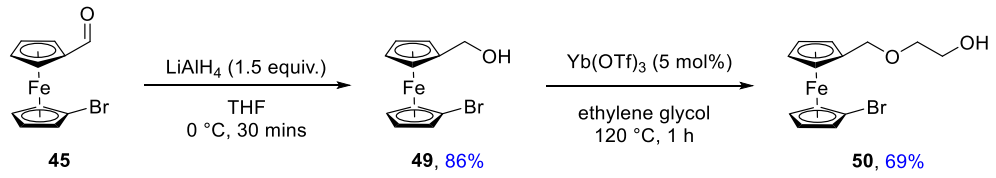
Scheme 2.25 Electrophilic substitution of ferrocenyl acetal

The telescoped chain extension was employed to extend the alkyl chain of these derivatives (*cf.* the synthesis of 3-ferrocenylpropan-1-ol **38**), however subjecting 1'-bromoferrocenecarboxaldehyde **45**, and the stannyl derivative **48** to this reaction resulted

in complete protodehalogenation and protodestannylation, and as such this route was deemed not suitable for substituted ferrocenes.

Due to the unsuitability of the Horner-Wadsworth-Emmons reaction, the acetals were converted to the corresponding alcohol. To increase the chain length, the pseudobenzylic alcohol can be subjected to the ytterbium (III) triflate catalysed etherification. This protocol was first applied to benzylic substrates,[151] where the ytterbium acts as a Lewis acid catalyst forming a stable benzylic carbocation, followed by nucleophilic substitution by excess diol. The reaction can be conducted on ferrocenyl methanols with varying lengths of diols.[134]

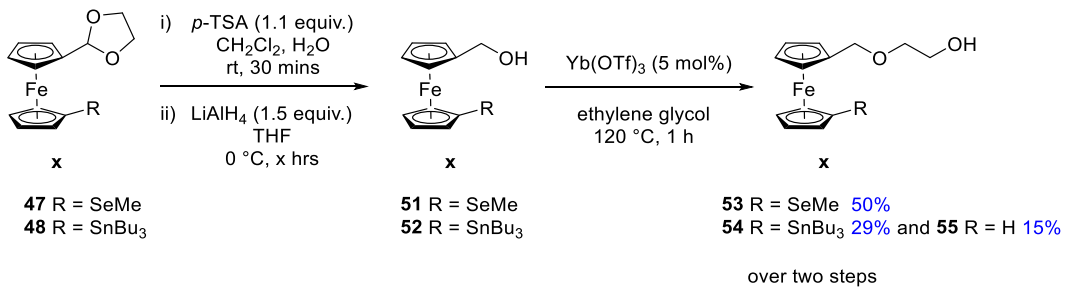
The reaction was applied to 1'-bromoferrocenecarboxaldehyde, first reducing the aldehyde with sodium borohydride, followed by treatment of the resultant alcohol with ytterbium (III) triflate (Scheme 2.26).



Scheme 2.26 Etherification of 1'-bromoferrocenecarboxaldehyde

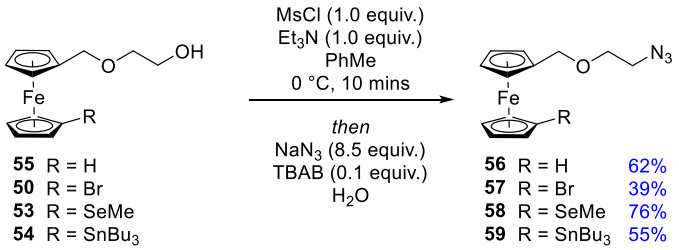
The substituted 1,3-dioxolanes were subjected to the same reaction, however upon deprotection it was found the aldehydes were particularly unstable, and as such the reaction was telescoped from the deprotection through to the chain extension (Scheme 2.27). Each step showed full consumption of starting material to a single spot by crude TLC, and as such was used without further purification at each stage.

It was found that the tributylstannane underwent a significant degree of protodestannylation during the telescoped procedure. A single spot was observed when carrying out TLC for the reduction to yield the intermediate alcohol **52**, however upon extension of the chain, the R_f of the stannyl and protio species separated, and as such it was possible to separate the two products by silica gel chromatography at the final step.



Scheme 2.27 Etherification of substituted ferrocenes

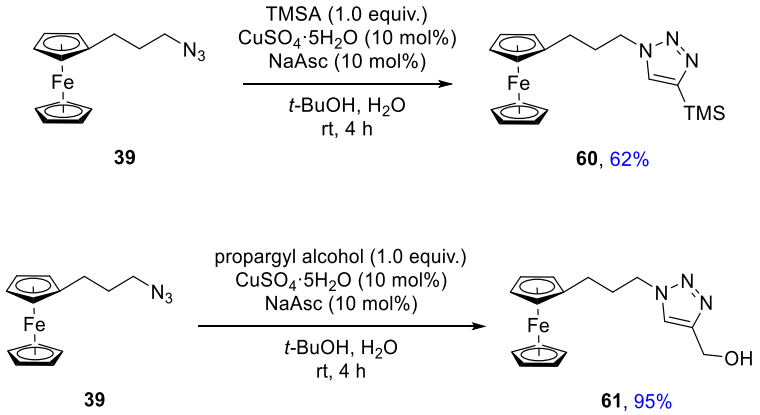
The conditions for conversion of the ferrocenyl alcohol to the azide developed previously were attempted on **53**, **54**, and **55**, however it was found that the conversions to the corresponding bromides *via* and Appel reaction were unsuccessful. These ferrocenyl alcohols were, however, successfully converted into the corresponding azides *via* a one-pot mesylation and nucleophilic substitution reaction (Scheme 2.28).



Scheme 2.28 Azidation of substituted ferrocenyl alcohols

2.3.2. Conjugation of ferrocene to 2'-deoxyuridine *via* the copper-catalysed azide-alkyne cycloaddition

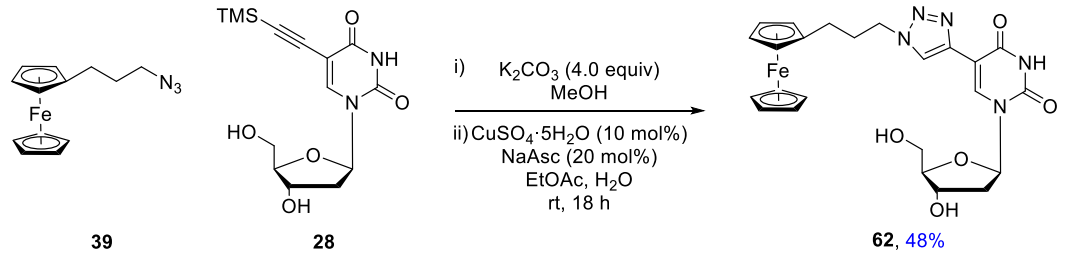
After the successful synthesis of a small library of ferrocenyl azides, the conjugation of these to the uridine was investigated. First, a test CuAAC reaction was carried out with simple alkynes to investigate the suitability of the ferrocenes with alkyne substrates (Scheme 2.29). According to literature procedures, the copper (II) sulfate and sodium ascorbate catalyst system was used, employing 10 mol% catalyst loading. Using 1-azido-3-ferrocenylpropane **39**, this was conjugated to both trimethylsilyl acetylene and propargyl alcohol with great success, showing that the ferrocene system is indeed stable under the conditions.



Scheme 2.29 Test CuAAC reactions with ferrocene

Having shown the ferrocene is stable under CuAAC conditions, it was then attempted to conjugate the ferrocene directly to the nucleoside. This was achieved by combining the ferrocene and the protected alkyne in the presence of copper (II) sulfate and sodium

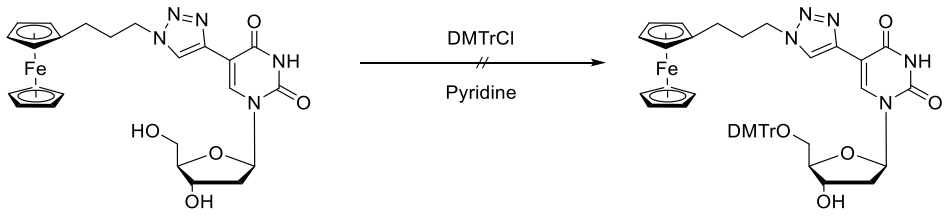
ascorbate. The reaction proceeded slowly, with the formation of a new spot occurring after only 2 hours. The spot continued to grow, and after stirring overnight at rt, the product was found to have precipitated out of solution to leave a yellow slurry. The reaction was subjected to a simple aqueous work-up to remove the salts and quick column to remove any unwanted starting materials remaining to generate the desired product in a 48% yield.



39 **28** **62**, 48%

Scheme 2.30 Conjugation of ferrocenyl azide to nucleoside

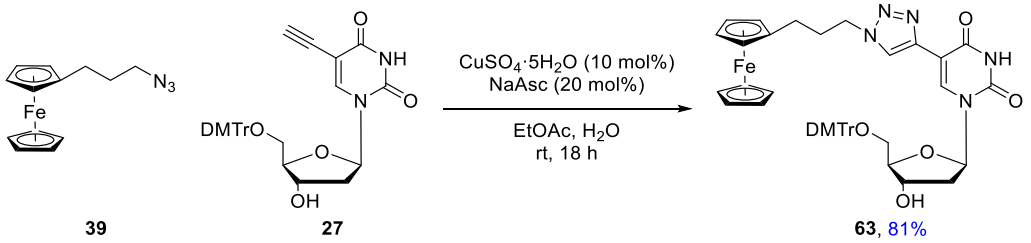
After the successful synthesis of the ferrocenylated deoxyuridine, it was then attempted to protect the 5'-hydroxyl group. Following standard literature procedures for tritylation, the nucleoside was treated with DMTrCl in pyridine. After 18 hours at room temperature, there was no indication of the formation of desired product, and crude ^{1}H NMR showed only starting materials.



62 **63**

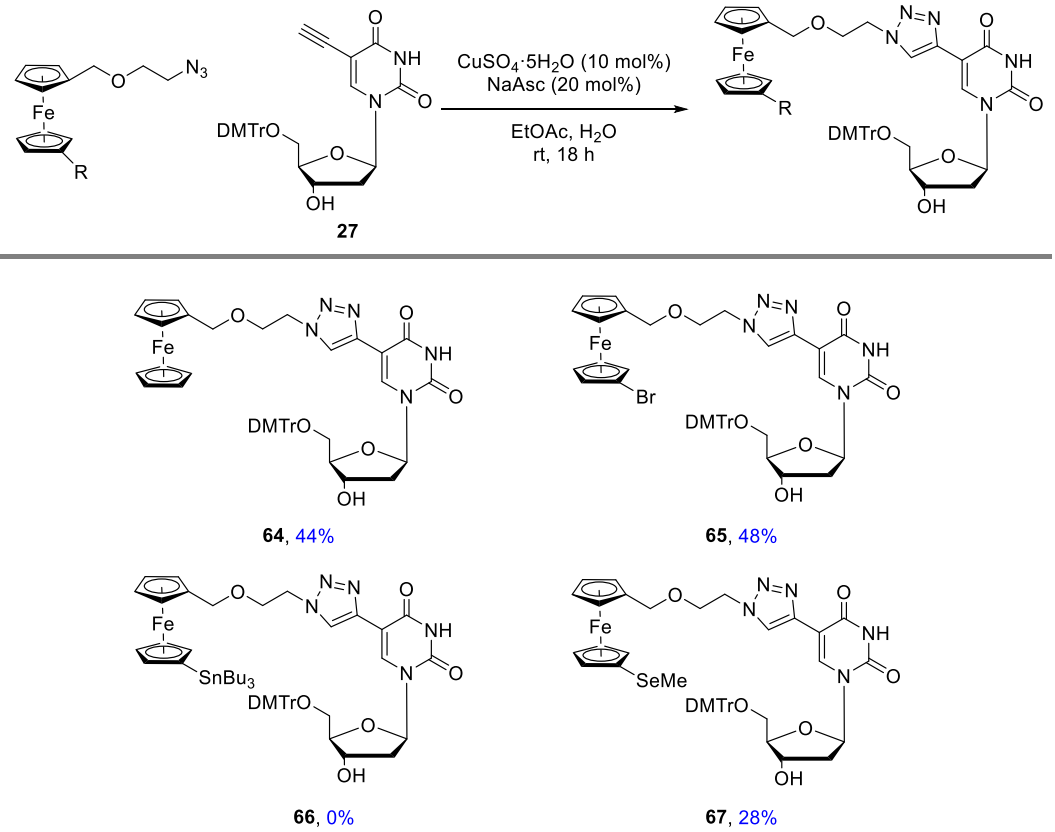
Scheme 2.31 Attempted tritylation of ferrocenyl nucleoside

DMTr protection occurs *via* an S_N1 reaction, leaving the stable benzylic carbocation intermediate. It is plausible that due to the bulky ferrocene in the 5-position of the uridine, this could block the large planar trityl cation and therefore no protection would occur. As such, the order of reactivity was changed, and it was decided to swap the protection and tritylation. Additionally, this would reduce the number of steps required after conjugation of the ferrocene. The DMTr protection was therefore carried out on the free alkyne substrate, and the CuAAC was then performed on the protected nucleoside. By carrying out the reaction in this order it was possible to synthesise the redox active protected nucleoside in good yields.



Scheme 2.32 CuAAC with pre-installed 5'-protecting group

The library of ferrocenylazides was applied to this reaction such that a range of ferrocenyl uridines were synthesised (Scheme 2.33).

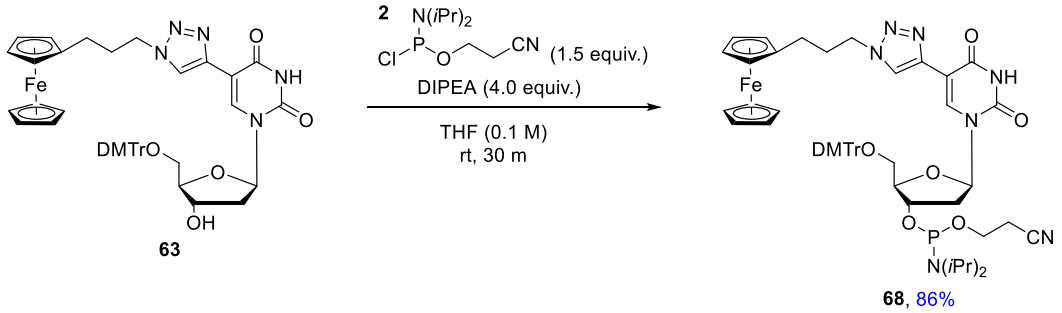


Scheme 2.33 Library of substituted ferrocenyl nucleosides

Under the CuAAC reaction conditions, the derivative **66** was not stable, instead yielding the protodestannylated species **64**. The yield of all reactions was disappointingly low at under 50% for all three, likely due to ferrocene decomposition, as evidenced by the brown colour of the aqueous layer upon work-up, and staining of the silica upon purification.

2.3.3. Phosphoramidite activation of redox active nucleosides

The final stage of chemical synthesis for the ferrocenyl uridines was to convert the 3'-hydroxyl into the phosphoramidite, ready for solid-phase oligonucleotide synthesis. This was attempted on the unsubstituted ferrocenyl nucleoside **63**.

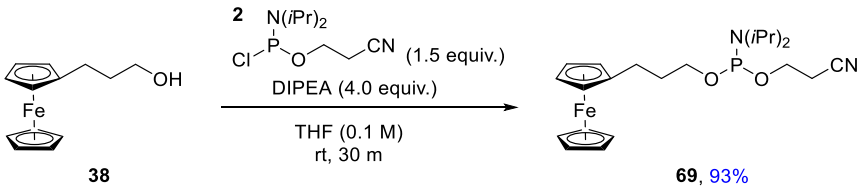


Scheme 2.34 Phosphoramidite formation of ferrocenyl 2'-deoxyuridine

The reaction proceeded smoothly, with indications of a successful reaction from DIPEA·HCl precipitation after only a few minutes, and a new spot forming by TLC after 10 minutes. Upon aqueous work-up and standard column chromatography, the new spot was isolated. Upon analysis, it was apparent that two products had been isolated. Diastereomers are to be expected as the molecule has multiple stereogenic centres, including a newly formed stereogenic phosphorous centre, which was formed with no stereocontrol. This would, in itself, lead to a complicated ^{1}H NMR as the diastereomers could not be separated in the eluent systems trialled, but the ^{1}H NMR spectrum was particularly crowded, and the ^{31}P NMR spectrum showed both a P(III) and a P(V) species, with the P(V) peak at 10 ppm being the major product. Phosphoramidites are known to readily oxidise to the phosphoramidate, and as such the reaction was repeated, however great care was taken to remove risk of oxidation.

All solvents used had argon bubbled through prior to use, including work-up and column solvents. The laboratory did not have air-free purification techniques, therefore the column was set up such that a nitrogen flow was used instead of traditional bellows, and secondary needle was attached to the spout of the column to ensure no additional oxygen entered the system. Additionally, as the compounds are coloured and the R_f of the starting material and product were significantly different, it was possible to collect directly into a round-bottom flask and evaporate solvents with minimum exposure to air. This level of care allowed the desired product to be isolated without any trace of the P(V) species.

As the ferrocenylthymidine phosphoramidite had been synthesised successfully, it was necessary to also synthesise a 5'-label of a similar size and substitution pattern to compare digestion properties of 5'-labelling *vs*. internal labelling. To this extent, it was chosen that 3-ferrocenylpropan-1-ol **38** was to be converted into the corresponding phosphoramidite **69**. This was achieved under the same conditions as the nucleoside phosphoramidite.



Scheme 2.35 Phosphoramidite formation of 5'- label

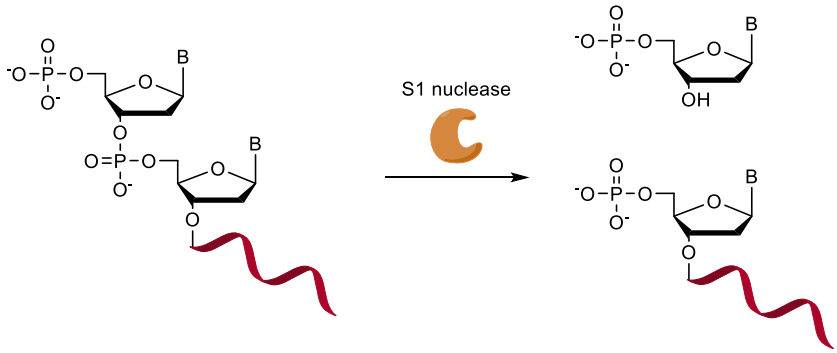
The non-nucleosidic phosphoramidites are significantly more stable than nucleoside derivatives, however caution was also taken to ensure the desired product was isolated as solely the P(III) species.

Due to low yields and small scale of the other ferrocenyl nucleosides **64**-**67**, they were not converted to the phosphoramidites.

2.4. S1 digestions of mono-ferrocene oligonucleotides

S1 Nuclease is a single-strand specific endonuclease, having activity in the 5'-3' direction. It cleaves DNA and RNA to release 5'-phosphorylated mono and oligonucleotides (Scheme 2.36).[152] S1 nuclease was chosen as a proof-of-concept digestion protocol, to investigate in a simple reaction whether the current could be increased. This would allow desired probe sequences to be synthesised and tested, without relying on PCR protocols or cDNA hybridisation.

The S1 assays have a lower level of complexity than double-stranded digestions as there are fewer components. This was used as a check to see whether the ferrocene attached to the oligonucleotide would inhibit enzymatic activity. If these probes did not digest under S1 nuclease conditions, then their applicability to a full assay would be in doubt.



Scheme 2.36 S1 nuclease activity

2.4.1. Internal Control sequence selection

The sequence chosen was based on Atlas Genetics' internal control (IC) probe sequence, which is used in their detection assays to confirm the result of the test. The IC sequence is a 23mer with 5 dT residues. It was decided to substitute the thymidine units for the modified ferrocenyluridine **68** in a sequential manner, from the 5'-end towards the 3'-end. This resulted in the design of six probes (Table 2.2). One probe was designed as a control (IC-C), to compare the internal labelling strategy to the 5'- labelling. The dT residues were then substituted for **68** in a stepwise manner.

The synthesis of these probes was conducted by ATDbio, who were provided with roughly 100 µmole of phosphoramidites **68** and **69** for the synthesis of the various probes. The probes were synthesised *via* the phosphoramidite method and purified and analysed *via* HPLC. HPLC-MS traces confirmed the successful inclusion of all ferrocene units (Appendix 5), however coupling efficiencies were not provided by the company.

Probe	Sequence
IC-C	5'Y-GCA CGA TCC CTT TCC TAA AGA CG-3'
IC-1	5'-[p]-GCA CGA U^{m144}CC CTT TCC TAA AGA CG-[p]-3'
IC-2	5'-[p]-GCA CGA U^{m144}CC CU^{m144}T TCC TAA AGA CG-[p]-3'
IC-3	5'-[p]-GCA CGA U^{m144}CC C$U^{m144}U^{m144}$ TCC TAA AGA CG-[p]-3'
IC-4	5'-[p]-GCA CGA U^{m144}CC C$U^{m144}U^{m144}$ U^{m144}CC TAA AGA CG-[p]-3'
IC-5	5'-[p]-GCA CGA U^{m144}CC C$U^{m144}U^{m144}$ U^{m144}CC U^{m144}AA AGA CG-[p]-3'

Table 2.2 Internal Control oligonucleotide sequences. [p] = phosphorylated.

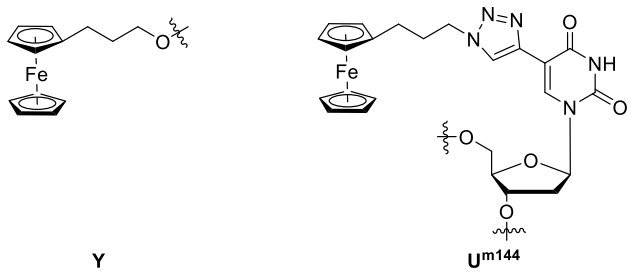


Figure 2.2 Key for Table 2.2

The synthesis of these probes was conducted by ATDbio, who were provided with roughly 100 μmole of phosphoramidite for each probe. HPLC-MS traces confirmed the successful inclusion of all ferrocene units, however coupling efficiencies were not provided by the company.

The multi-labelled probes were then subjected to S1 nuclease digestion to investigate their possible increased signal. It was decided not to attempt to optimise the reaction buffer, and as such all S1 digestions were conducted according to the supplier's protocol.[153]

The background oxidation of the probes was first investigated (Figure 2.3). When comparing background oxidations, the 5'-labelled probe IC-C results in a higher signal than the mono internally labelled probe IC-1. When the ferrocene building block is incorporated on the 5'-end, it is hypothesised it can access the electrode surface more readily, hence leading to oxidation at the surface. When this is moved internal to the oligonucleotide, this is not as easily accessible due to formation of secondary structures, therefore a comparatively reduced current.

This effect is not continued throughout, however, as substitution for additional ferrocene units leads to an increase in current. The effective concentration of ferrocene is increasing and as such leads to an increase in current. The additional ferrocene units could also disrupt the formation of secondary structures, resulting in the ferrocene cores being more available. There is correlation between number of ferrocenes and current generated, such that the

background rate of the penta-labelled probe IC-5 has the highest background. The increase in current for the internally labelled probes exhibits a linear relationship.

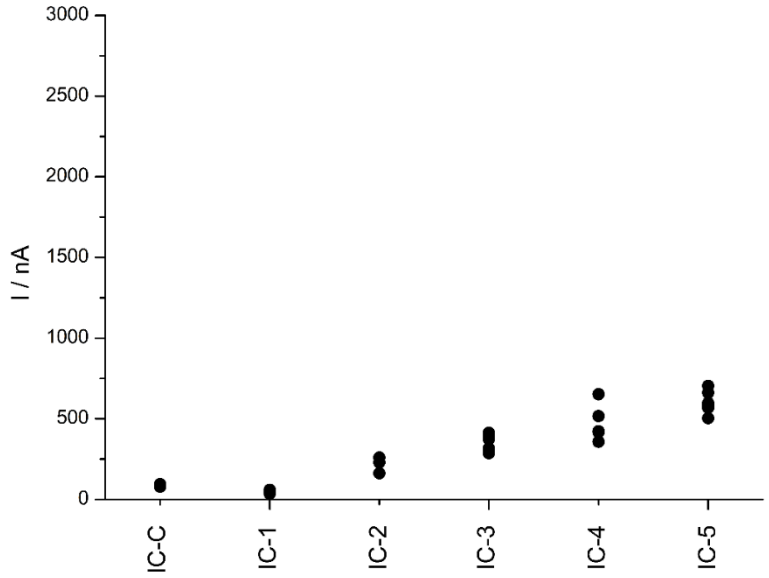


Figure 2.3 IC probes (6 μM), pH 4.5 NaOAc buffer, 37 °C, 30 minutes.

The digestion of the probes was then investigated to assess their applicability in a diagnostic assay. It is believed the intact probe is too large to generate a significant level of current, as the size of the probes means limited diffusion will occur and upon introduction of S1 nuclease the 3'- *O-P* bond is cleaved therefore realising smaller nucleosides, which are able to diffuse to the electrode and undergo redox cycling.

Initially, the probes were studied by DPV after treatment with S1 nuclease for 15 minutes. The IC-C, IC-1, and IC-2 probes showed full digestion in 15 minutes at 37 °C. This is clearly seen in the voltammograms as both an increase in current and a shift in E_{ox} occurs. For the IC-C probe, the background E_{ox} is centred around 190 mV, which shifts to 100 mV upon digestion (Figure 2.4).

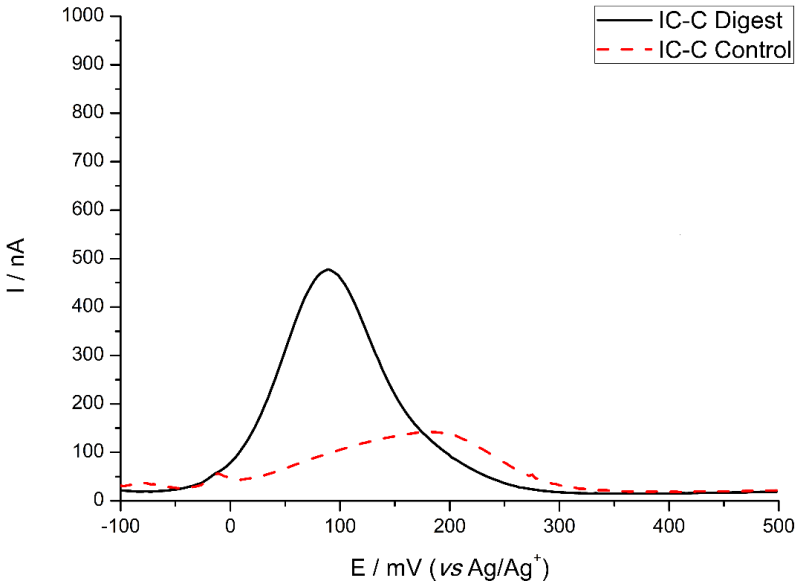


Figure 2.4 IC-C, S1 nuclease (10 U), pH 4.5 NaOAc buffer, 37 °C, 15 minutes

All internally labelled probes exhibited a similar shift in oxidation potential, such that the intact oligonucleotide E_{ox} is centred around 260 mV, shifting to 140 mV upon digestion (Figure 2.5).

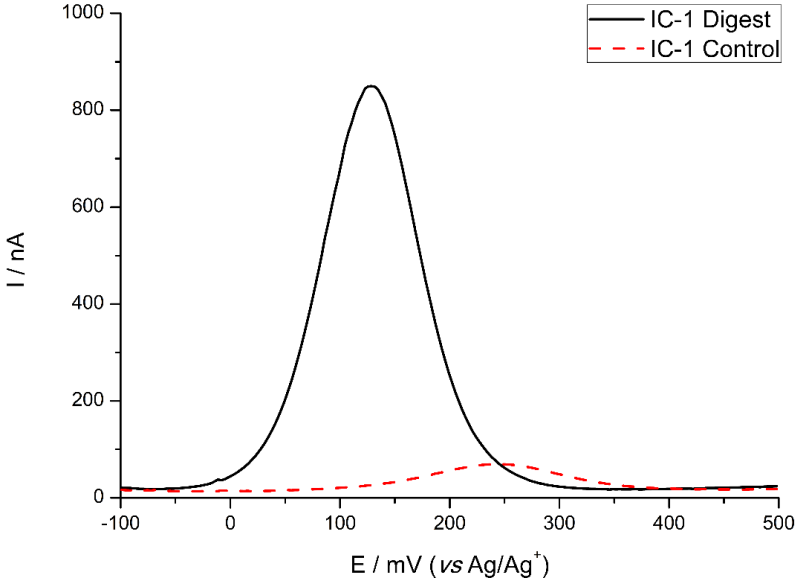


Figure 2.5 IC-1, S1 nuclease (10 U), pH 4.5 NaOAc buffer, 37 °C, 15 minutes

The E_{ox} of all the internally labelled probes comes at the same potential for the digested probes indicating that the same species is being oxidised. The incorporation of five ferrocene units has an effect on the current generated, but the electronic environment of the different units remains constant, or at least very similar, in all probes.

Probes IC-3, IC-4 and IC-5 did not digest within 15 minutes. This was observed on the voltammograms, where the spectrum is no longer a single peak. This can be attributed to the intact probe as the secondary peak, observed as a shoulder, comes at the same oxidation potential as the background (Figure 2.6).

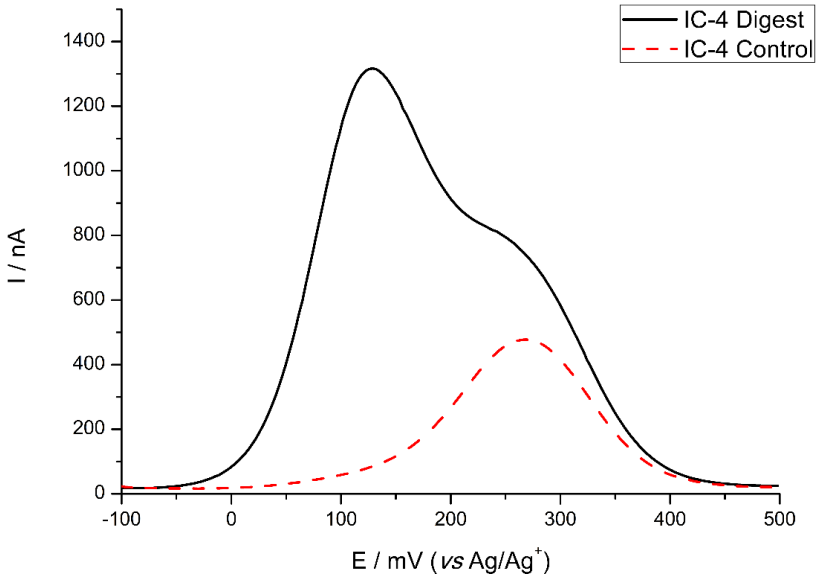


Figure 2.6 IC-4, S1 nuclease (10 U), pH 4.5 NaOAc buffer, 37 °C, 15 minutes

The digestion time was increased to 30 minutes at 37 °C for all probes to allow for full digestion. This allowed the IC-3, IC-4, and IC-5 probes to digest, confirmed by the absence of a secondary peak, to the extent that only a single probe was observed in the voltammogram.

Having established suitable digestion conditions for all probes, the experiments were repeated to obtain a spread of data. Due to the variable nature of the screen-printed electrodes, the value from one scan to another can vary significantly, and as such the runs must be repeated to obtain a reliable data range (Figure 2.7).

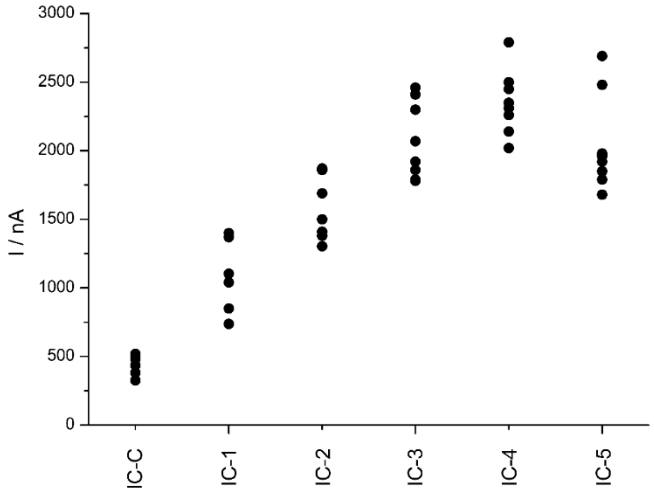


Figure 2.7 IC probes (6 μM), pH 4.5 NaOAc buffer, S1 nuclease (10 U). 37 °C, 30 minutes

Upon digestion the internally labelled probe IC-1 resulted in an increased signal compared to the 5'-labelled probe IC-C. This is hypothesised to be mainly due to increased solubility of the probe in the aqueous medium, as the probe is conjugated directly to the nucleobase which will improve its solubility. Introducing multiple ferrocenyl units resulted in an increase for two and three ferrocenyl units, however no significant increase in signal was observed when four or five units were incorporated.

Theoretically, the maximum signal that could be generated for IC-5 is 5 times the current of the IC-1, as if full digestion has occurred the concentration will be 5 times as high. To investigate this further, IC-1 was studied at 30 μM to determine whether the higher labelled probes were inhibiting enzyme activity, or the high ferrocene concentration could have caused an electrode overload (Figure 2.8).

The background rates were studied at the higher concentration of 30 μM. Increasing the background concentration did not result in a fivefold increase in background. Upon digestion, the increased concentration of IC-1 resulted in a large increase in signal. This indicated that there is no overload occurring at the electrode, and that the lower signal can be attributed to reduced digestion. From the voltammograms, it is clear that there was no undigested probe present, as there was no secondary peak or shoulder. This suggests the probe is still being fully digested despite the increased substitution.

If the digested products are smaller, *i.e.* mono or di nucleotides, it will result in a higher signal. It is possible that the size of the products digested is different, resulting in the lower signal. S1 nuclease may digest the probe to a suitable degree to lower its oxidation potential such that only a single peak is observed, however leave the probe large enough so it is unable to diffuse to the electrode, therefore generating a lower current.

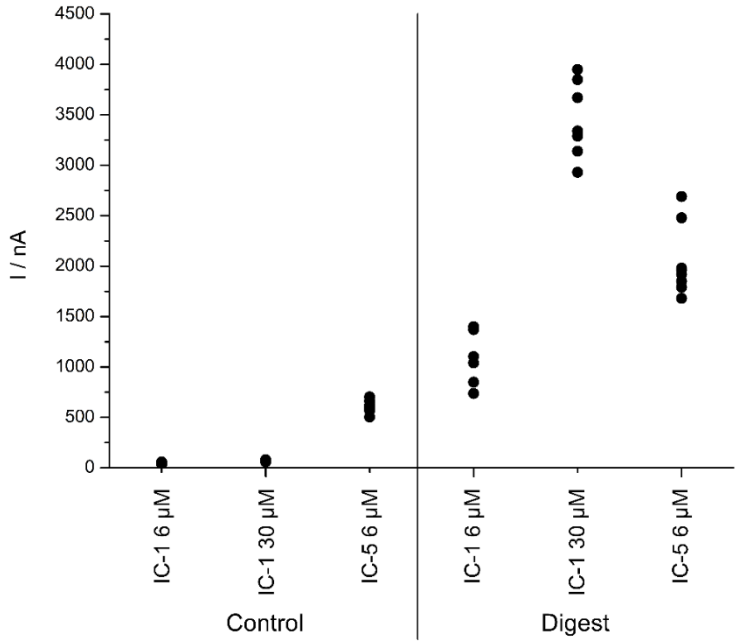


Figure 2.8 IC internally labelled concentration studies IC-1 (6 µM and 30 µM) and IC-5 (6 µM), S1 nuclease (10 U), pH 4.5 NaOAc buffer, 37 °C, 30 minutes.

It was hypothesised that there is no real benefit from having more than three labels on the IC sequence. The increased background rate, but lack of increase in digested signal, results in a lower sensitivity for the probes and as such reduces the utility of the highly substituted probes.

2.4.2. CT sequence selection

It was postulated that the proximity of the substituted thymidine units to each other could be inhibiting the enzymatic activity leading to lower signal of the probes than what was expected. Due to the proximity of the thymidine residues in the IC sequence, it was difficult to spread the labels out and as such, it was desirable to move to a sequence in which the U^m could be separated. As such, the *C. trachomatis* (CT) sequence used by Atlas Genetics in their assays was chosen as a suitable probe.[154] As such, six more probes were designed and synthesised, this time focussing on increasing the number of nucleotides between each label.

Probe	Sequence
CT-mC	5'Y - CTG TCC GCT GGT TCT TCC TTA CT-[p]-3'
CT-1	5'-[p]-CU^{m144}G TCC GCT GGT TCT TCC TTA CT-[p]-3'
CT-2	5'-[p]-CU^{m144}G TCC GCU^{m144} GGT TCT TCC TTA CT-[p]-3'
CT-3	5'-[p]-CU^{m144}G TCC GCU^{m144} GGT TCT TCC TTA CU^{m}-[p]-3'
CT-4	5'-[p]-CU^{m144}G U^{m144}CC GCU^{m144} GGT TCT TCC TTA CU^{m}-[p]-3'
CT-5	5'-[p]-CU^{m144}G U^{m144}CC GCU^{m144} GGT TCT TCC U^{m144}TA CU^{m144}-[p]-3'

Table 2.3 *Chlamydia trachomatis* oligonucleotide sequences. [p] = phosphorylated.

These probes were subjected to the same digestion protocol with S1 nuclease (Figure 2.9). The background rate of these probes follows a similar trend to the IC equivalents, showing a slight decrease in signal for the mono internally labelled probe CT-1 *vs.* CT-C, and then a linear increase in signal for each additional ferrocene unit thereafter.

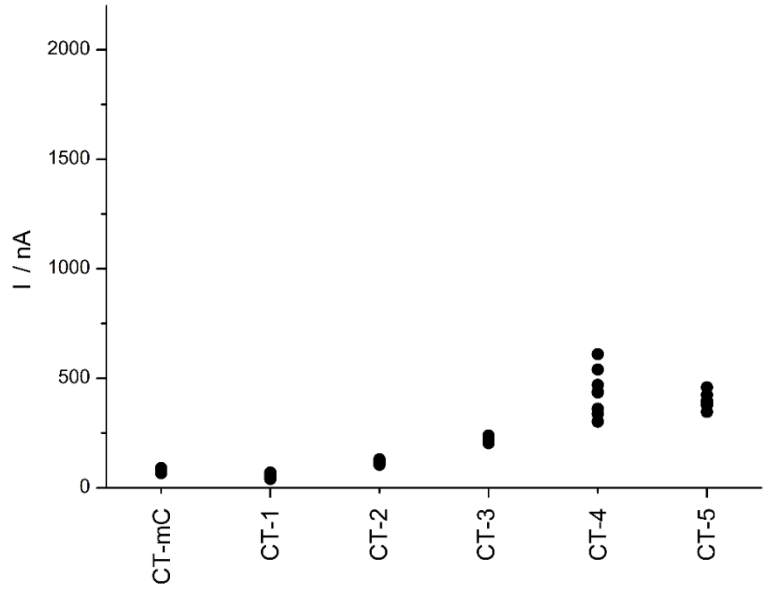


Figure 2.9 CT probes (6 μM), pH 4.5 NaOAc buffer, 37 °C, 30 minutes

Upon digestion, the mono internally labelled probe CT-1 resulted in a greater signal than the control probe CT-C, with an increase in signal also being observed for the CT-2 probe, however similar to findings with the IC probes, the introduction of 3 or more ferrocenes did not result in any significant increase in signal (Figure 2.10).

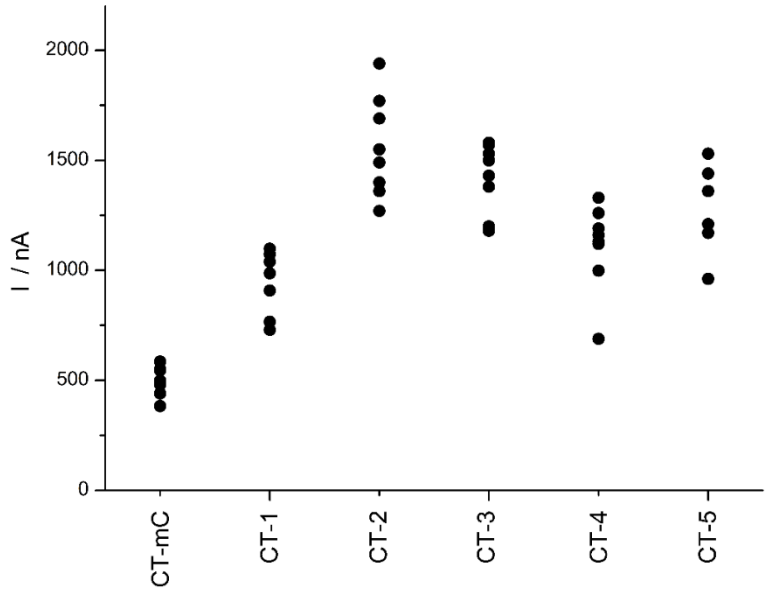


Figure 2.10 CT probes (6 μM), pH 4.5 NaOAc buffer, S1 nuclease (10 U). 37 °C, 30 minutes

The CT sequence gave lower signals than the IC sequence. The max current generated was 2000 nA, however the IC sequence could give currents up to 2500 nA. Additionally, the

signal plateau was observed at a lower substitution value of two units, (*cf.* three for IC sequence). This was contrary to initial thoughts, as it was hypothesised that by increasing spacing between functionalised dT units would improve digestion properties, however this was not the case.

Probes CT-1 to CT-5 were studied by mass spectrometry (MS) to study the digestion profile *via* UPLC-MS, however due to the already low concentration of the probes, no information fragments or intact probes could be identified. Due to the small scale and low yields of conjugating the MS-labelled species to uridine, it was deemed unviable to convert these into the corresponding phosphoramidites.

2.5. Conclusion

A range of ferrocenyl azides were coupled to 2'-deoxyuridine utilising the CuAAC reaction. One of these redox active nucleobases was chosen as a suitable test substrate in which to incorporate into oligonucleotides. The internal control and *C. trachomatis* sequences used by Atlas Genetics were chosen and up to five ferrocenes were successfully incorporated into the 23mer oligonucleotides as confirmed by HPLC-MS for all probes. The oligonucleotides showed a linear increase in background current for the substitution level. It was shown that it was possible to increase the signal for single strand S1 digestions, indicating compatibility with enzymes and the probes were shown to undergo digestion despite high level of substitution, however there was no benefit to including more than three ferrocenes on a single strand. The continued digestion of all probes was promising for further development into systems which could be used for the detection of pathogenic DNA *via* double stranded digestions.

3. Design, synthesis, and testing of di-ferrocenylated nucleosides

3.1. Introduction

Chapter 2 focussed on the synthesis of a library of mono-ferrocenyl azides, with initial studies investigating the suitability of internally labelled oligonucleotides in solution-based DNA detection. It was shown that upon digestion, the new labelling strategy allowed for increased signal in a simple single-stranded S1 digestion.

It was next investigated whether the signal could be further improved in a digestion assay through increasing the number of ferrocenes, without increased substitution of thymidine units, as a signal plateau was observed and no increase in signal could be generated past three substituted nucleotides.

There were two main goals for the work in this chapter, namely increasing the signal that could be generated in a digestion and generating a system which would enable improved multiplex detection.

Atlas Genetics has found great success with di-ferrocene labels (Figure 3.1), finding that they give higher signals and tend to have smaller peak widths. As well as improved sensitivity, the smaller peak width allows increased multiplex capability, as this enables a greater number of ferrocenes with different oxidation potentials to be incorporated into a single voltammogram.[155] However, only labels **70** and **71** have been successfully incorporated into diagnostic assays due to limited sensitivity of mono-labels and difficulties with the synthesis of other labels.

One of the first labels used by Atlas Genetics in a diagnostic assay was di-275 (**70**), a di-substituted amine which contained two ferrocene cores. This label, and derivatives thereof (Figure 3.1), gave a greater than twofold increase when used in diagnostic assays, indicating a synergistic effect of the multiple redox centres. The ferrocene label nomenclature used relates to both the structure and oxidation potential *e.g*. "di-275" (**70**) is a bis(ferrocenylmethyl) amine, whose oxidation potential is at 275 mV in pH 9.0 Trizma buffer.

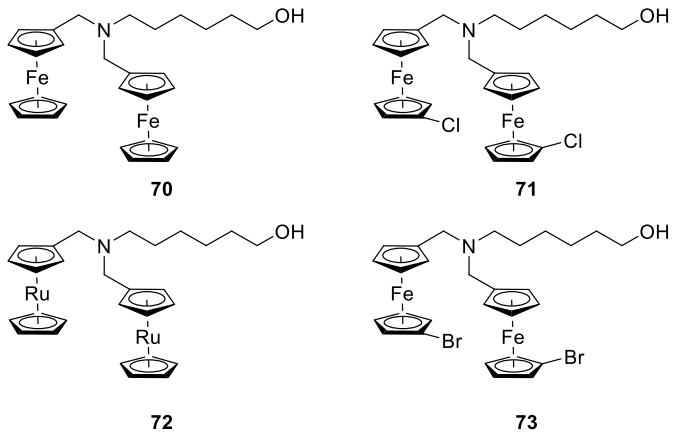


Figure 3.1 Atlas Genetics' di-ferrocene labels

To date, Atlas Genetics have not carried out any investigations into the crystal structures of the di-ferrocenes. It was hypothesised that the presence of the tertiary amine and shape of the di-ferrocenes could be responsible for the vastly improved electronic properties. There are scant literature examples of this bis(ferrocenylmethyl) amine structure, with Kaifer and co-workers dominating the examples with their work towards dendritic ferrocenes.[112,156–161] The bis(ferrocenylmethyl) amine crystal structures refined by Tice *et al.*[160] (**74**), Alvarez *et al.*[159] (**75**), and Kerr *et al.*[161] (**76**) showed an interesting perpendicular arrangement between the ferrocene cores, while the structure of tris(ferrocenylmethyl) amine (**77**) reported by Sun *et al.*[157] also exhibited the interesting arrangement to some degree.

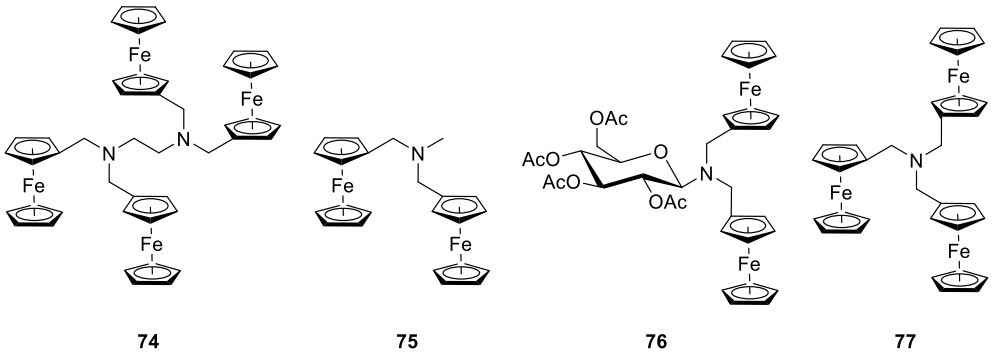
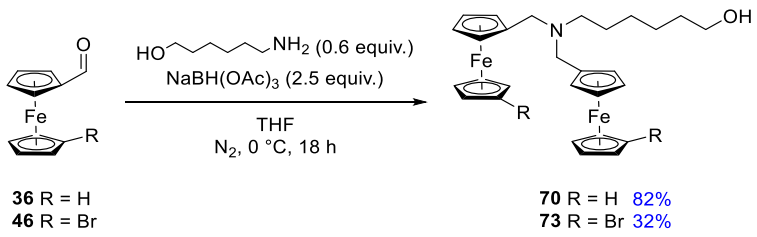


Figure 3.2 bis and tris (ferrocenylmethyl) amines exhibiting perpendicular arrangement

It was of interest to investigate the structure of the library of Atlas Genetics' labels to determine whether they also adopt a similar spatial arrangement to those previously reported. As such, crystals of a selection of bis(metallocenylmethyl) amine structures were grown. Previous attempts to grow crystals have proven futile, with crystals being twinned or amorphous and therefore unsuitable for crystallography. Samples of **70** and **73** were synthesised (Scheme 3.1), while the chloride **71** and ruthenocene **72** derivatives were

donated by Atlas Genetics, however **71** is unsuitable for crystallography as it is an oil at room temperature.



Scheme 3.1 Reductive amination of ferrocenecarboxaldehydes

To grow crystals of suitable quality, the compounds were first subjected to standard recrystallisation from hot ethyl acetate to improve sample purity after silica gel chromatography. Single crystals were then grown through combined warming and layering methods, first dissolving in the minimum warm ethyl acetate, followed by gentle layering of petroleum ether (60–80 °C) on top. This was a suitable method and single crystals of each compound were grown.

It was interesting to discover a perpendicular orientation between the two metallocene centres, which was observed for all three metallocenes studied. The most pronounced arrangement was that of the 1'-bromo derivative **73** (Figure 3.4), while the ruthenocene derivative did not exhibit such an obvious structure (Figure 3.5). It was hypothesised that the orientation between the two ferrocene units would allow for communication between the two, resulting in a synergistic effect and as such giving a greater than twofold increase in signal. The structures of these labels are similar to the structures previously reported in the literature, which could give further evidence towards the origin of the improved performance of these structures.

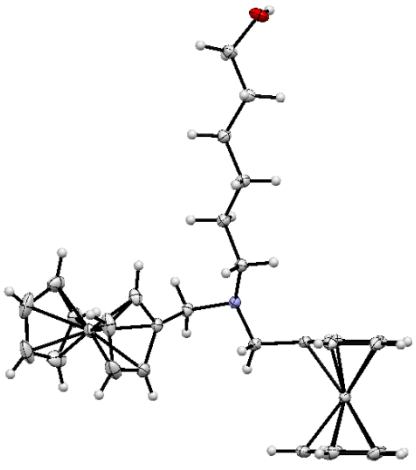


Figure 3.3 ORTEP diagram of di-275 (**70**)

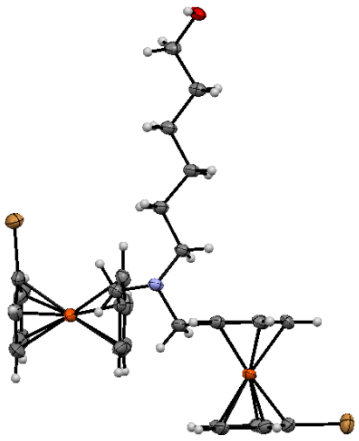


Figure 3.4 ORTEP diagram of di-437 (**73**)

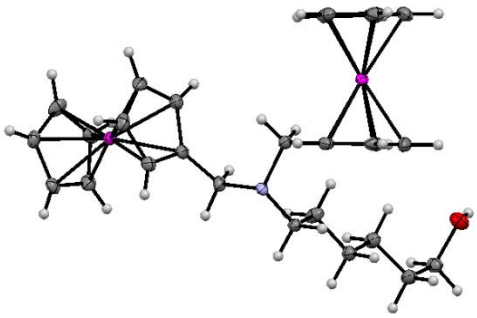


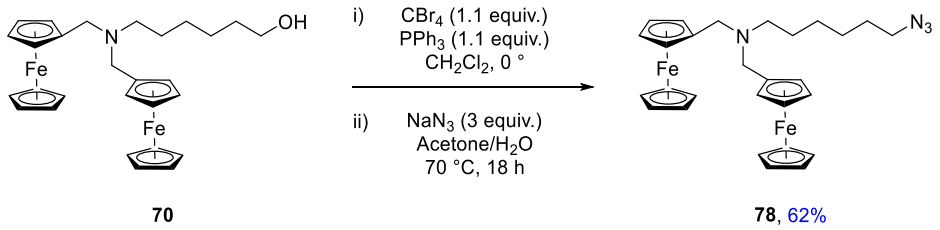
Figure 3.5 ORTEP diagram of di-763 (**72**)

Knowledge of the relationship between structure and enzymatic performance could allow for informed and intelligent design of probes, selecting probes whose signals are far superior. These crystal structures could form the basis of computational studies in which the relationship is further explored.

3.2. Di-ferrocene synthesis and nucleobase conjugation

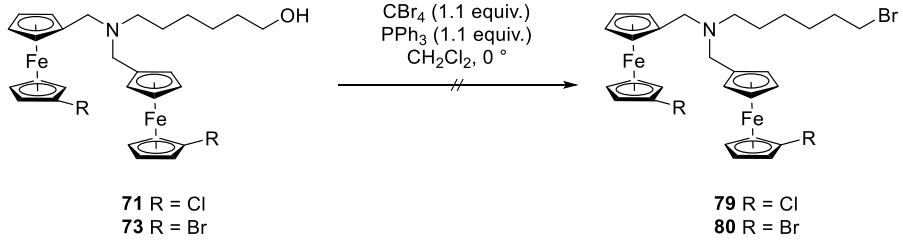
Given the superior performance of di-ferrocenes in diagnostic assays, it was desirable to incorporate the di-ferrocene labels into the internal labelling strategy. As such, the synthesis of a library of di-ferrocenyl labels was attempted, first synthesising the ferrocenyl alcohols, followed by conversion to the corresponding azide.

Following the success of the route towards 3-azido-1-ferrocenylpropane **39**, it was attempted generate the di-ferrocenyl azide **78** *via* a telescoped Appel reaction and nucleophilic substitution with sodium azide (Scheme 3.2). Pleasingly, this gave the desired product in a 62% yield over two steps.



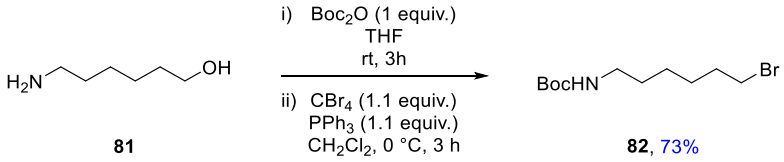
Scheme 3.2 Azidation of di-275

It was also attempted to convert alcohols **71** and **73** to the corresponding azides *via* the Appel reaction, however attempts to convert the labels to the bromide failed, returning only starting material.



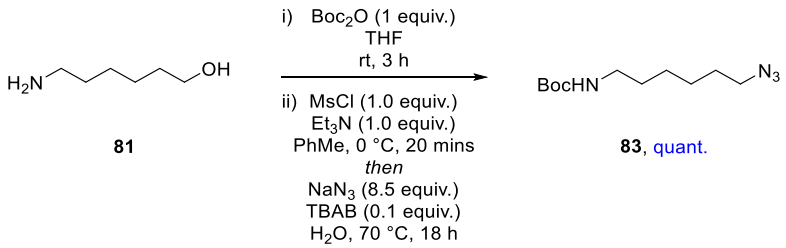
Scheme 3.3 Appel reaction of substituted di-ferrocenes

It was thought that by installing the azide functionality before the reductive amination could overcome this issue, improving synthesis by first forming the azide from 6-aminohexan-1-ol (**81**). The amine was first Boc protected using standard conditions, and formation of the bromide was achieved *via* an Appel reaction to give the desired product **82**.



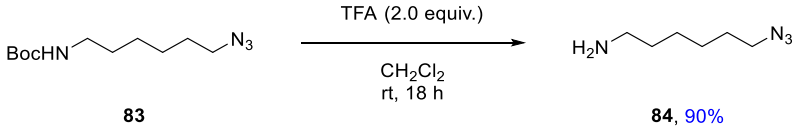
Scheme 3.4 Protection and Appel reaction of 6-aminohexan-1-ol

After finding success for the conversion of primary bromides to azides with sodium azide in acetone, **82** was subjected to the same conditions, however the reaction did not proceed, and no formation of product was observed. As the azide nucleophile failed to displace the bromide, further optimisation was performed to convert the alcohol to an azide. It was found that by forming the mesylate intermediate from **81**, followed by addition of azide *in situ* allowed formation of the desired intermediate protected compound **83** in quantitative yield (Scheme 3.5).



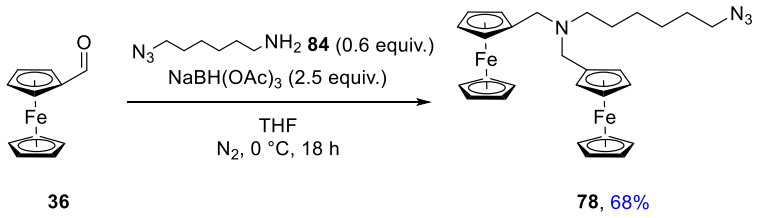
Scheme 3.5 Azidation of 6-aminohexan-1-ol

This was then deprotected using trifluoroacetic acid, followed by a basic aqueous work up to yield the desired aminoazide **84** (Scheme 3.6).



Scheme 3.6 TFA deprotection of Boc-aminoazide

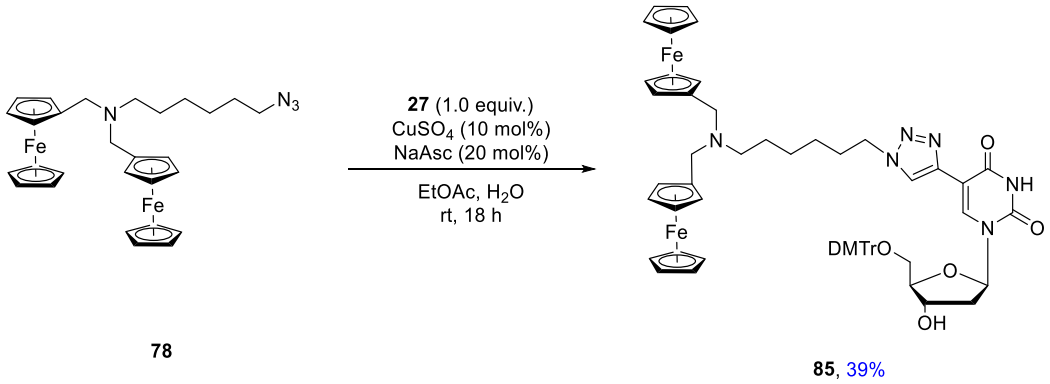
The aminoazide **84** was then substituted into the reductive amination, and this allowed the reaction to proceed cleanly, giving the desired product in 68% yield (Scheme 3.7).



Scheme 3.7 Reductive amination with aminoazide

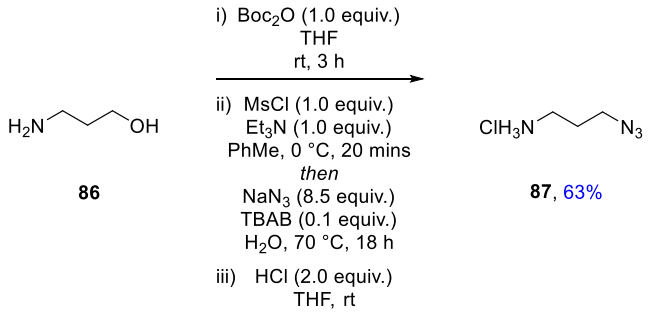
By performing the reactions in this order, this led to a 17% increase in yield from ferrocenecarboxaldehyde. Additionally, this would allow for reduced complexity when synthesising substituted di-ferrocenes due to the fewer linear steps.

To test the route towards redox active nucleobases, the ferrocene **78** was conjugated to 2'-deoxyuridine *via* the CuAAC conditions developed in Chapter 2 (Scheme 3.8), which was isolated in a low yield of 39%, despite full consumption of starting material observed through TLC.



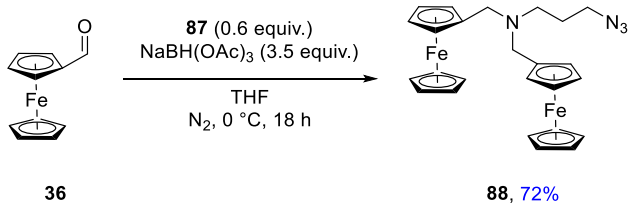
Scheme 3.8 CuAAC reaction between 5-ethynyl-2'-deoxyuridine and di-ferrocenylazide.

Due to the size of the additional ferrocene unit and the lengthy 6-carbon linker, it was hypothesised that a shorter linker should be synthesised to reduce the structural flexibility in the system as this could interfere with duplex formation or inhibit enzymatic digestion of the probes. To this extent, 3-aminopropan-1-ol **86** was converted to the azide *via* the same method as the hexanol derivative. The reaction was telescoped from the alcohol through to the HCl salt of the azide **27**, obtaining the desired product **87** in a 63% yield (Scheme 3.9). The salt is incredibly hygroscopic and difficult to work with unless fully dry, which could not be achieved *via* high vacuum oil pump alone, however a vacuum desiccator with P_2O_5 was found to efficiently dry the compound to a constant mass after one week.



Scheme 3.9 Conversion of alcohol to azide

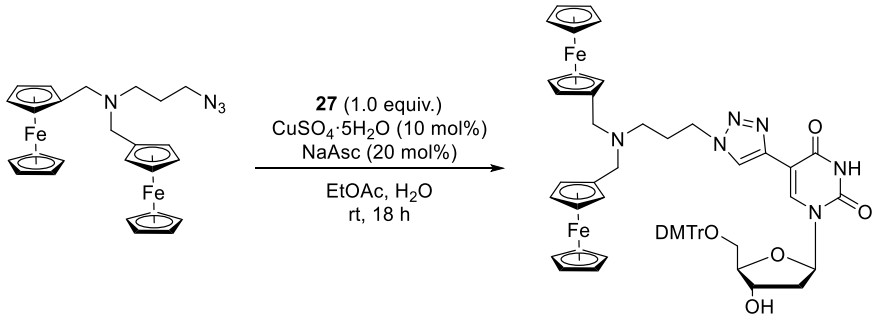
This shorter linker **87** was employed in the reductive amination of ferrocenecarboxaldehyde (Scheme 3.10). Pleasingly, the formation of the freebase *in situ* did not hinder the reaction, generating the desired product **88** in 72%, similar to the yield of the 6-carbon derivative (68%).



36 88, 72%

Scheme 3.10 Reductive amination with shorter aminoazide

It was then sought to conjugate the ferrocenyl azide **88** deoxyuridine, which was achieved through treatment of the azide and alkyne with the copper (II) sulfate and sodium ascorbate catalyst system, in the same manner as the mono-ferrocene derivatives, yielding the desired product **89** in a 58% yield (Scheme 3.11).



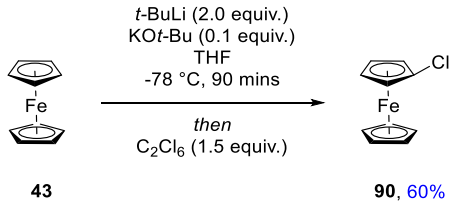
88 89, 65%

Scheme 3.11 Click between shorter di-ferrocenylazide and 5-ethynyl-2'-deoxyuridine.

After the successful synthesis of an unsubstituted di-ferrocenyl deoxyuridine, attention was turned to synthesise substituted ferrocene derivatives. As the mono-ferrocene library focussed on the synthesis of mass spectrometry labels, it was of interest to generate a system with multiplex capability to allow for the detection of multiple targets in a single assay utilising the internal labelling strategy. As such, electron-rich and electron-poor derivatives were adapted from Atlas Genetics' patents. Having already synthesised the unsubstituted di-275 derivative, it was required to synthesise electron-poor and electron-rich ferrocenecarboxaldehydes to enable the formation of substituted di-ferrocenes *via* the reductive amination reaction.

The halogen series offers an opportunity to increase electron-withdrawing capability with relatively simple synthetic ease. As the synthetic route towards substituted ferrocenes requires access to 1'-bromoferrocenecarboxaldehyde **45** (synthesis in Chapter 2), this gives access to an electron-poor substituent with no additional synthesis required.

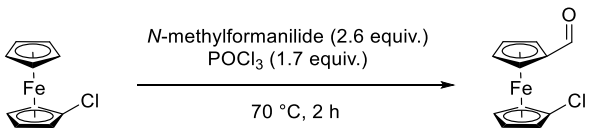
To keep in line with AG's library, the chloride derivative was also synthesised. The synthesis of chloroferrocene **90** was achieved *via* the use of Schlosser's base to achieve mono-lithiated ferrocene (Scheme 3.12).[162] Lithioferrocene was then quenched with hexachloroethane, and subjected to oxidative purification, *cf.* 1,1'-dibromoferrocene,[150] to remove any unreacted ferrocene. The product was recrystallised to yield the chloroferrocene in 60% yield.



43 **90**, 60%

Scheme 3.12 Chlorination of ferrocene

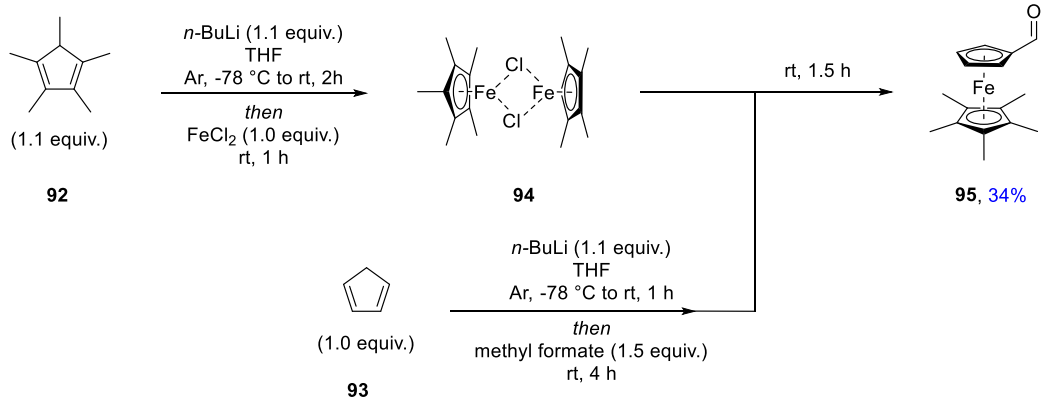
This was then converted to the corresponding carboxaldehyde *via* the modified Vilsmeier-Haack reaction as a melt to install the desired aldehyde functionality (Scheme 3.13). The reaction proceeded with 4:1 regioselectivity, with the 1,1'- substituted carboxaldehyde being the major product. As the unsubstituted cyclopentadienyl ring is more electron rich, the regioselectivity generated is in favour of the top ring, confirmed by the splitting pattern upon ^{1}H NMR analysis.



90 **91**, 27%

Scheme 3.13 Formylation of chloroferrocene

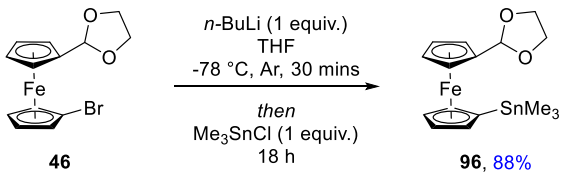
An electron-rich ferrocenylaldehyde was also required, and to achieve this, 1',2',3',4',5'-pentamethylferrocenecarboxaldehyde **95** was synthesised. The ferrocene core was constructed from iron (II) chloride, Cp, and Cp* with the aldehyde pre-installed (Scheme 3.14).[155]



Scheme 3.14 Synthesis of 1',2',3',4',5'-pentamethylferrocenecarboxaldehyde

Although the yield of the reaction is not particularly high (34%), and has complex timings, the reaction is simple to purify. Unsubstituted ferrocenes are very apolar, and as such any mixed sandwich would be difficult to purify as it is inevitable that both decamethylferrocene and ferrocene would also form in the reaction. By pre-installing the aldehyde moiety on the cyclopentadienyl ring, any mixed sandwich that forms will have significantly different R_f values, therefore allowing the material to be isolated cleanly *via* silica gel chromatography without the need for sublimation techniques.

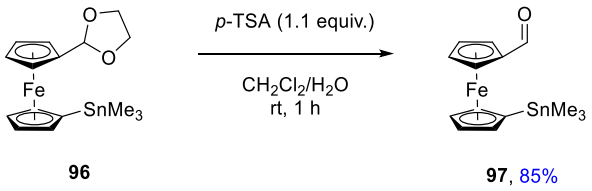
The introduction of a mass spectrometry substituent was also investigated. Earlier studies with the tributylstannane derivative proved unsuccessful due to high levels of protodestannylation observed during synthesis of the azide, and regardless of the low yields, the stannane did not survive the conditions of the CuAAC reaction, instead yielding the protodestannylated ferrocene derivative (Chapter 2). It was thought that by moving from tributyl to trimethyl derivative, the stannane may be more stable and therefore survive the CuAAC reaction to yield a stannylated deoxyuridine. To this extent, 1'-bromoferrocene-1,3-dioxolane **46** was converted to the trimethylstannane derivative **96**, using trimethyltin chloride as a source of electrophilic tin.



Scheme 3.15 Lithium halogen exchange and quench with tin electrophile

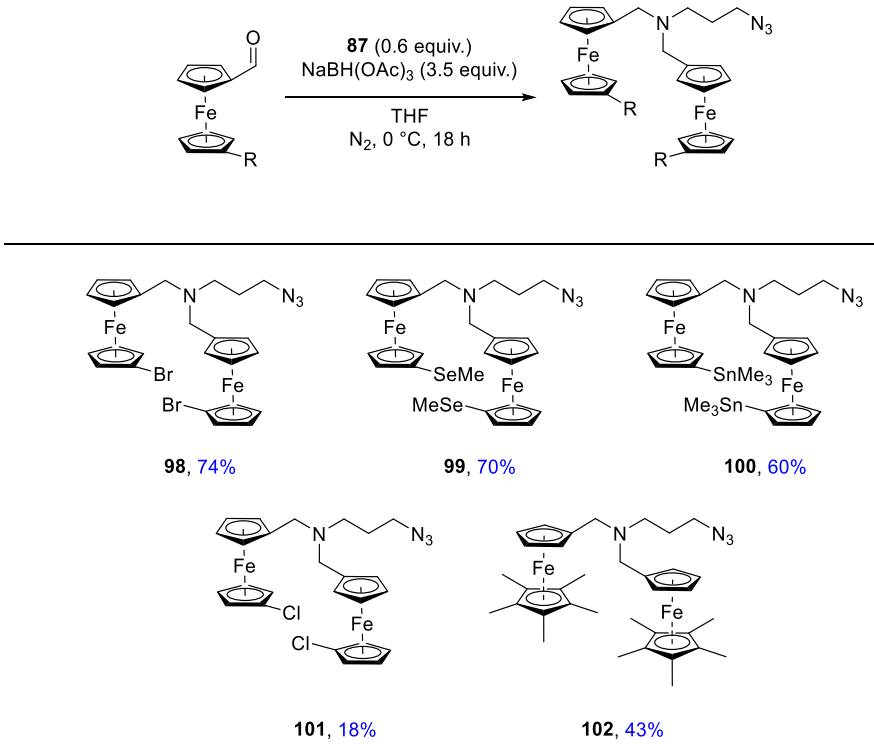
This was then converted into the corresponding aldehyde **97** through treatment of the acetal with an excess of *p*-toluenesulfonic acid in dichloromethane and water. The reaction proceeds cleanly and is high yielding with no trace of the protio species, however as is

common with substituted ferrocenecarboxaldehydes, the resulting aldehyde **97** was unstable and as such must be used immediately or stored at –20 °C over argon.



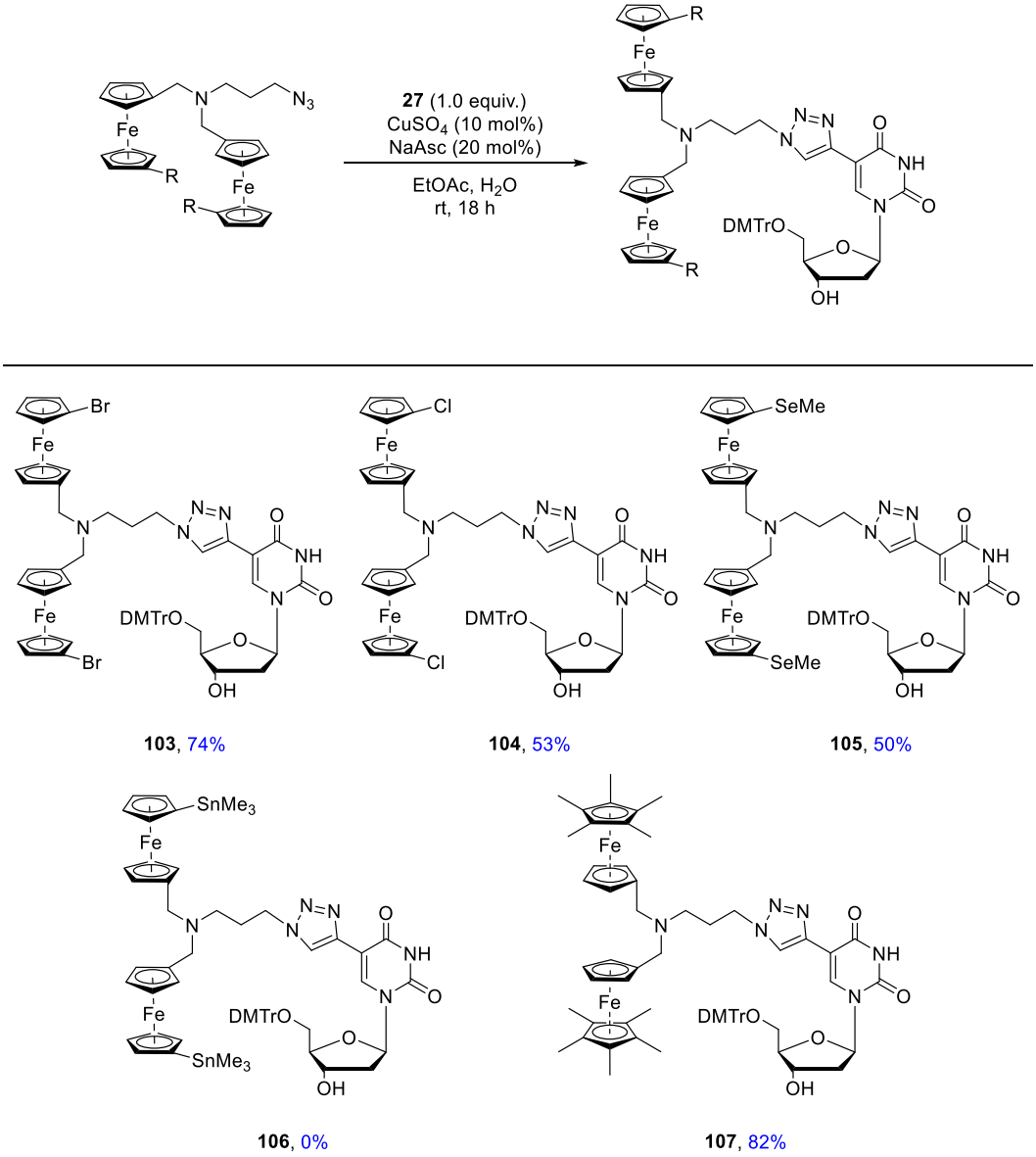
Scheme 3.16 Acetal deprotection of stannylated ferrocene

The reductive amination was then performed on the library of ferrocenylcarboxaldehydes that had been synthesised. The reaction was tolerant of all substituents, yielding the di-ferrocenyl azides in 18–74% yields (Scheme 3.17), and gave minimal levels of the undesired alcohol by-product. The stannyl derivative was well tolerated in the reaction, giving the desired product **100** in 60% yield, however both the chloride **101** and pentamethyl **102** derivatives performed poorly. This was clear during purification as both derivatives cause staining of silica, indicative of ferrocene decomposition. The chloride derivative left significant levels of brown residue on the silica, indicative of oxidation from Fe(II) to Fe(III), whereas the pentamethyl variant left green staining, suggesting the formation of a different Fe(II) complex.



Scheme 3.17 Reductive aminations on ferrocenecarboxaldehydes

The CuAAC reactions were performed on 5-ethynyl-2'-deoxyuridine with the di-ferrocene library (Scheme 3.18). The unsubstituted, bromide, chloride, and pentamethyl derivatives were well tolerated in the reaction, all giving yields above 50%, however the stannylated derivative **106** was not isolated, instead showing complete protodestannylation during the reaction.



Scheme 3.18 Click reactions of substituted ferrocenes and deoxyuridine

Previously, Atlas Genetics have used the unsubstituted and chloro di-ferrocenes as a duplex detection method, however they have struggled to find a third label of suitable sensitivity. The pentamethyl derivative has never been fully evaluated by Atlas Genetics due to poor initial screening results, so it was of interest to investigate whether this could be improved through internal labelling.

To assess the multiplex capability, compounds **89**, **103**, and **107** were dissolved in pH 9.0 Tris buffer, and subjected to voltammetric analysis. The three labels in a single, equimolar solution gave three easily distinguishable peaks, with only a slight overlap in the baseline (Figure 3.6).

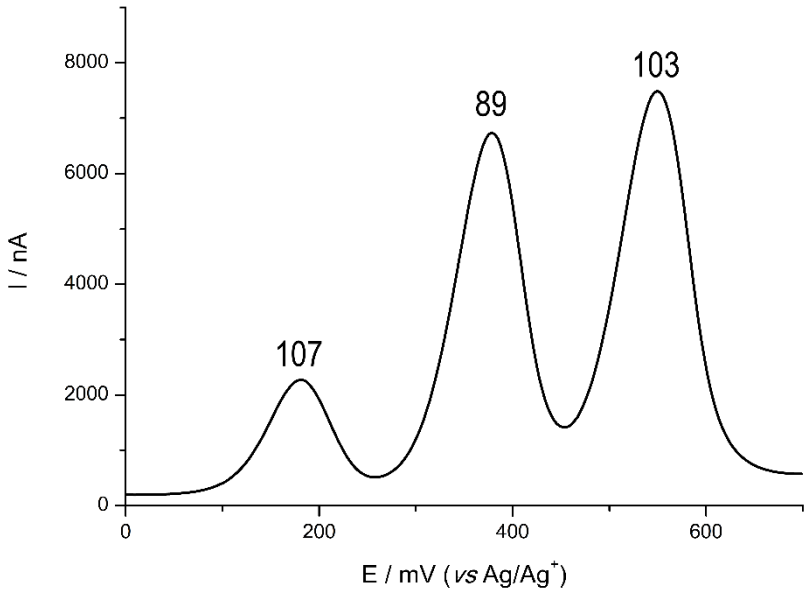
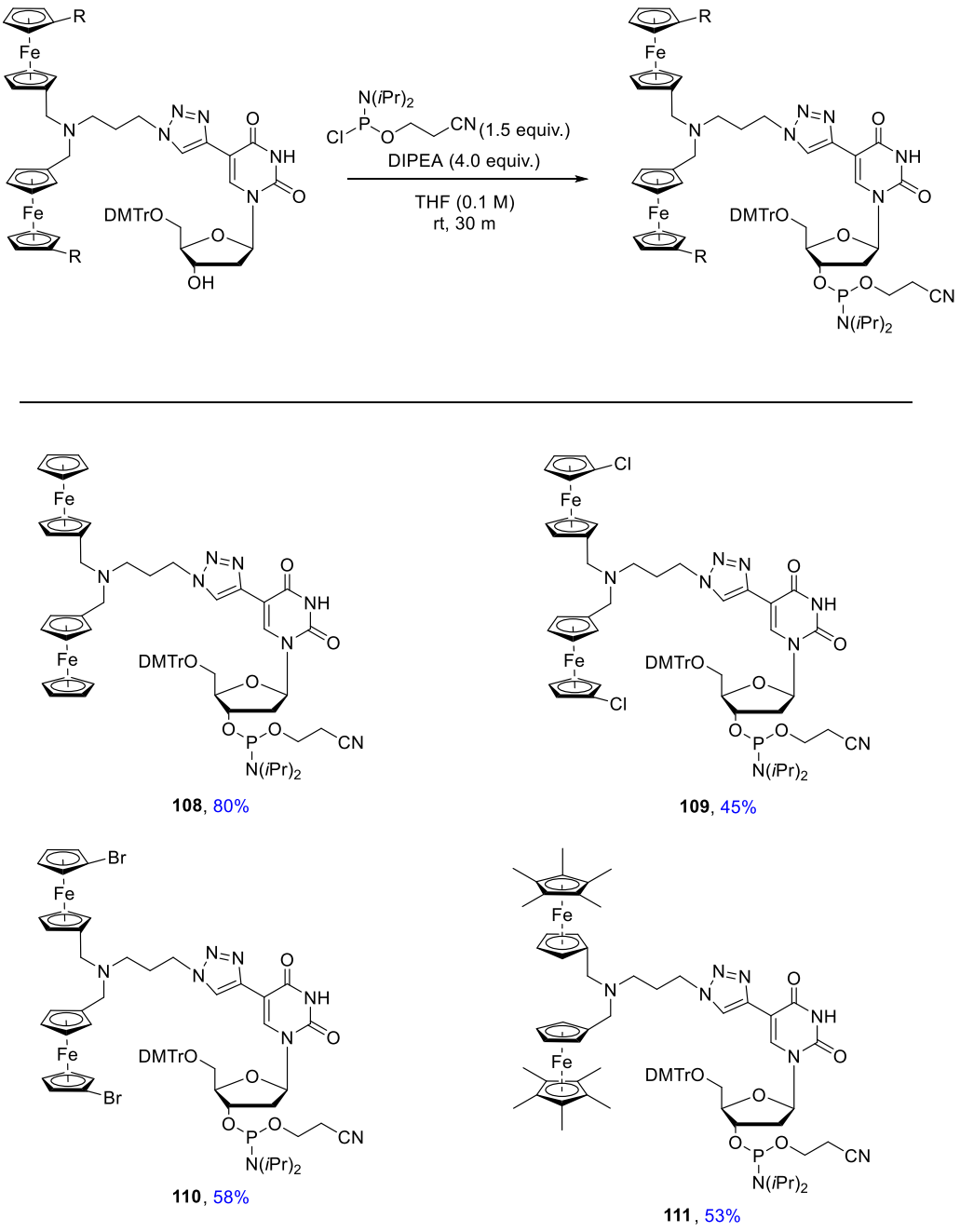


Figure 3.6 DPV of Multiplex labels **89**, **103**, **107**, 5 μM, PCR buffer

It was then sought to synthesise the phosphoramidites of each label. In each case, the reaction proceeds cleanly, with early indications of success visible by the precipitation of DIPEA·HCl. The formation of the phosphoramidites proceeds in 45–80% yields (Scheme 3.19).



Scheme 3.19 Phosphoramidite formation of di-ferrocenyldeoxyuridines

The reaction can be tracked *via* TLC without the need for visualisation methods, with the reaction showing the formation of a single new spot after 10 minutes. As was the case with the mono-ferrocenes, the R_f of the phosphoramidite is significantly higher than the free 3'-OH, allowing for visual isolation of desired product during chromatography.

A pragmatic approach was employed for the isolation of desired product, opting for purity rather than yield. Due to the compounds' susceptibility to oxidation when solvated, TLCs are not used to identify desired column fragments, instead utilising the compounds' colour to isolate desired fractions. As such, to minimise risk of including starting material in the isolated product, some of the desired product remains on the column.

3.3. S1 digestions of di-ferrocene oligonucleotides

The mono-ferrocene probes gave promising results, however it was clear that there was a limit in the current that could be generated due to the increased substitution of the probes. It was therefore investigated whether an increase in effective ferrocene concentration could be achieved *via* the introduction of the di-ferrocenes, rather than increase the level of substitution on the oligonucleotide probes. In theory, this would allow for a twofold increase in signal for the same level of oligonucleotide substitution.

The *C. trachomatis* sequence was chosen to allow future incorporation in a double stranded DNA digestion assay with clinical targets. Atlas Genetics have had great success with the label whose core **108** is based on, and as such this nucleoside was chosen for incorporation into the CT sequence (Table 3.1). All labels were shown to have been fully incorporated into the probe sequences confirmed by HPLC-MS provided by the oligonucleotide suppliers.

Probe	Sequence
CT-dC	5'-**Z**- CTG TCC GCT GGT TCT TCC TTA CT-[p]-3'
CT-6	5'-[p]-C$\underline{U^{d275}}$G TCC GCT GGT TCT TCC TTA CT-[p]-3'
CT-7	5'-[p]-C$\underline{U^{d275}}$G TCC GC$\underline{U^{275}}$ GGT TCT TCC TTA CT-[p]-3'
CT-8	5'-[p]-C$\underline{U^{d275}}$G TCC GC$\underline{U^{275}}$ GGT TCT TCC TTA C$\underline{U^{d275}}$-[p]-3'
CT-9	5'-[p]-C$\underline{U^{d275}}$G $\underline{U^{d275}}$CC GC$\underline{U^{d275}}$ GGT TCT TCC TTA C$\underline{U^{d275}}$-[p]-3'
CT-10	5'-[p]-C$\underline{U^{d275}}$G $\underline{U^{d275}}$CC GC$\underline{U^{d275}}$ GGT TCT TCC $\underline{U^{d275}}$TA C$\underline{U^{d275}}$-[p]-3'

Table 3.1 *Chlamydia trachomatis* probe di-ferrocene sequences. [p] = phosphorylated.

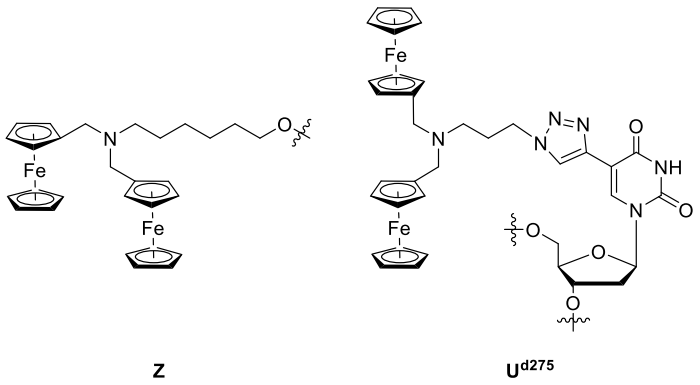


Figure 3.7 Key for Table 3.1

The digestions were carried out at the reduced concentration of 3 μM (previously 6 μM for mono-ferrocenes), to account for the increased effective concentration of ferrocene compared to the mono labels. The probes were digested using the same protocol as the mono-ferrocene labels discussed in Chapter 2. Upon treatment with S1 nuclease, the control 5'-ferrocenylated probe CT-dC was shown to undergo full digestion in 30 minutes, and the

digested product resulted in a shift in E_{ox} of roughly 50 mV when compared to the intact oligonucleotide (Figure 3.8).

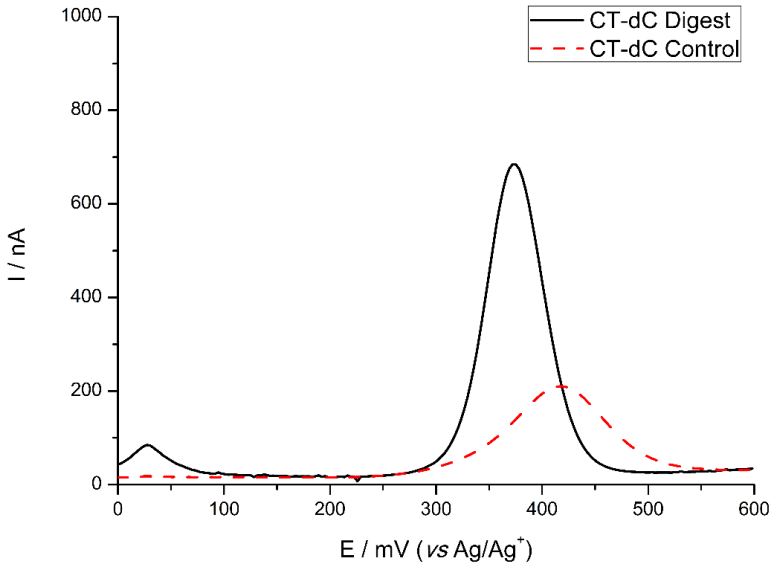


Figure 3.8 DPV of intact oligonucleotide CT-dC against digested probe. pH 4.5 NaOAc buffer, S1 nuclease (10 U), 37 °C, 30 mins

The internally labelled probes were also subjected to digestion under the same conditions, and it was shown all probes undergo full digestion in 30 minutes. The shift in oxidation potential between the intact and digested probe was similar at roughly 60 mV (Figure 3.9).

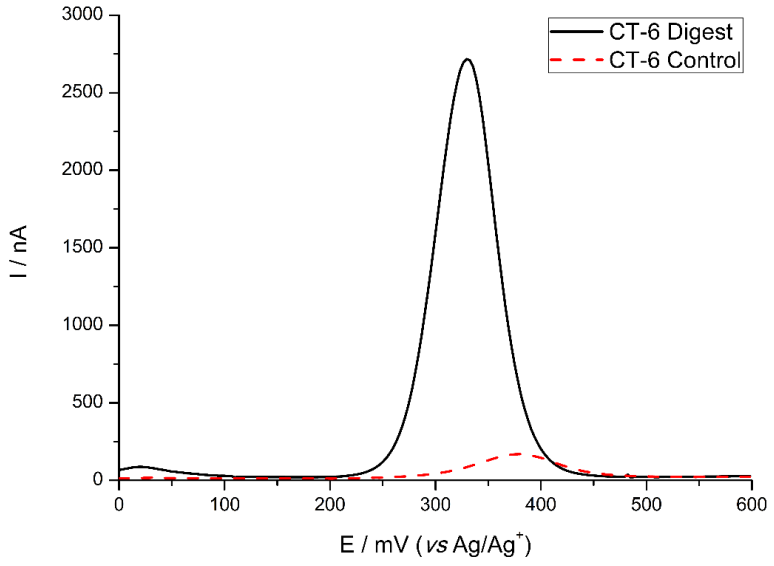


Figure 3.9 DPV of intact oligonucleotide CT-6 against digested probe. pH 4.5 NaOAc buffer, S1 nuclease (10 U), 37 °C, 30 mins

The shift in oxidation potential observed was much smaller than that of the mono-ferrocenes. This is likely due to the proximity of the labels to the nucleoside. In the mono-ferrocene derivatives, the linker between the triazole and the ferrocene moiety was much smaller, whereas with the di-ferrocenyl labels, they have two additional atoms between the ferrocene and the triazole. In theory, this makes it more difficult to determine whether the probes have undergone full digestion. In the case of the mono-ferrocenylated probes, the E_{ox} of the two species have distinct oxidation potentials, and it is clear when there is a second peak. As the digested and undigested E_{ox} of the di-ferrocene probes are close together, the presence of a shoulder would potentially not be as clear as in previous digestion studies, as it would be masked by the large peak height of the digested product.

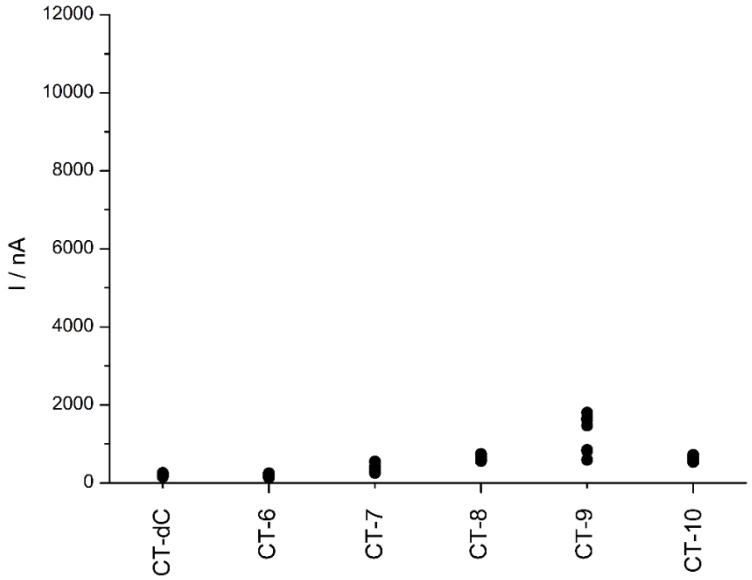


Figure 3.10 Di-ferrocene CT probe background, pH 3.4 NaOAc buffer, 30 mins, 37 °C

As was observed in Chapter 2 for both IC and CT sequences, moving from 5'- to internal labelling resulted in a slight reduction in background signal for CT-6 *vs.* CT-dC, and a linear increase in background oxidation of the probes CT-6 through CT-10 (Figure 3.10).

The probes proved significantly more sensitive than their mono-ferrocene equivalents. For the mono-ferrocene oligonucleotide probes, neither IC nor CT sequences were able to generate currents above 3000 nA at 6 μM. With the di-ferrocene CT probes, it was possible to increase this signal up to 10000 nA at only 3 μM, a huge increase in current compared to previous oligonucleotides (Figure 3.11).

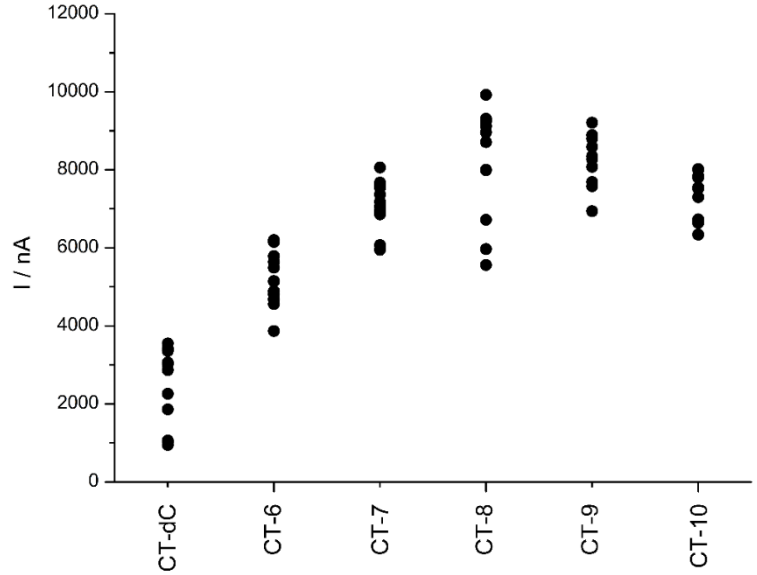


Figure 3.11 di275-CT probes, S1 nuclease, 37 °C, 30 minutes, Autolab detection

The digestion appears to follow a similar trend to the mono-ferrocene probes (Chapter 2), with a maximum current being generated with the incorporation of two labels. As most of the data points overlap for 2, 3, 4, and 5 substitutions, it can be said that no clear benefit arises from having more than two labels. With the mono-ferrocene label in Chapter 2, a signal increase was observed up to the point of three substitutions, however with the di-ferrocene this maximum was observed at two. The only difference in the probes is the label used, as such it is likely that there is some degree of enzyme inhibition caused by the bulky di-ferrocene.

3.3.1. Comparison of probe sensitivity

To quantify any improvements between traditional 5'-labelling and internally labelled probes, the sensitivity and limit of detection (LOD) were calculated in the range of 0.5 μM to 6.0 μM (Table 3.2).

Probe	Sensitivity ($nA\mu M^{-1}$)	Limit of Detection (μM)
CT-mC	76	0.128
CT-1	149	0.074
CT-dC	808	0.005
CT-6	1452	0.003

Table 3.2 Sensitivity and LOD calculations for selected probes

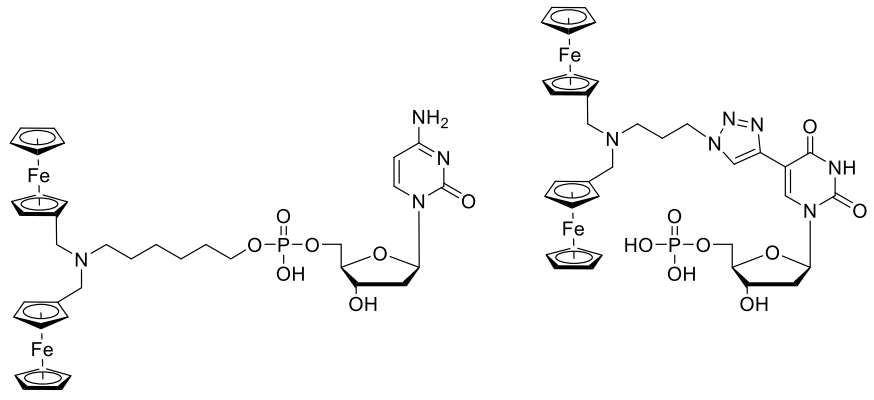
Internal labelling resulted in a twofold increase in sensitivity for both the mono-ferrocenes and the di-ferrocenes. The di-ferrocene labels are known to perform better in diagnostic

assays, however have never been evaluated for their sensitivity. These studies indicate the di-ferrocene labels gave a tenfold improvement in signal compared to the mono-ferrocene equivalent.

3.3.2. Mass spectrometry analysis

It is believed the increase in signal from the control to the digestion can be attributed to the release of mononucleotides which are able to diffuse to the electrode surface. The intact 23mer is unable to diffuse to the electrode surface, but the small nucleotide has increased availability at the electrode, therefore leading to higher electrochemical signals.

To confirm this hypothesis, the digestion mixtures were subjected to HRMS, searching for monomers, dimers, and trimers of the digest products. Probes CT-dC and CT-6 were subjected to digestion, identifying the $[M-H]^-$ peak in both for the expected mononucleotide released. CT-dC is conjugated to the 5'-end where cytidine is the terminal nucleotide, therefore releasing the 3'-hydroxyl nucleotide into solution (**112**), while for CT-6 the presence of the phosphorylated ferrocenyl-uridine CuAAC adduct (**113**) was also confirmed (Appendix 6). No di- or tri-nucleotide species were detected in either solution.



112 **113**

Figure 3.12 Structures of mononucleotides released by S1 nuclease

The control digestions were also analysed, displaying the intact oligonucleotide for both probes. The monomers were not detected in either of the control solutions.

3.4. Conclusion

The internal labelling strategy was expanded to di-ferrocene structures to study the potential improved signal through increased ferrocene concentration. A range of di-ferrocenyluridine phosphoramidites were synthesised, allowing for multiplex capability. One of these labels was incorporated into a clinically relevant 23mer showing improved signal over the mono-ferrocene probes developed in Chapter 2.

4. Internal Labelling in Diagnostic Assays

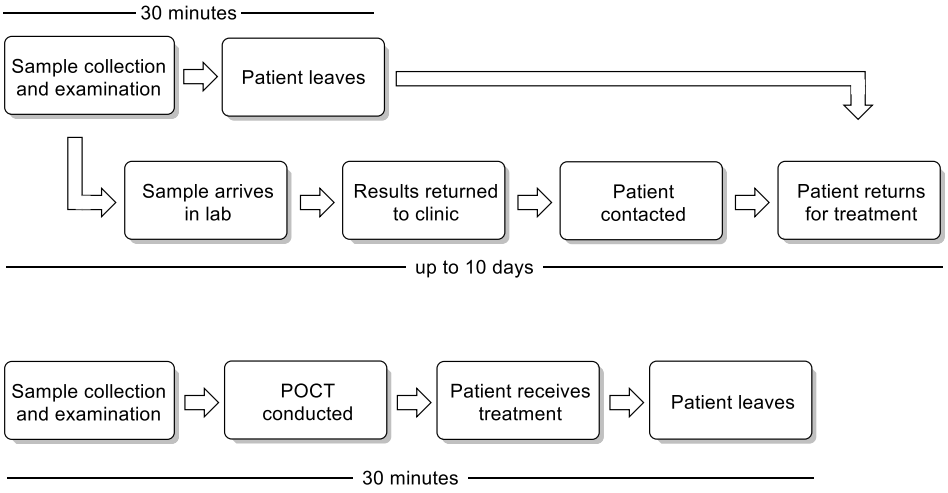
4.1. Introduction

Sexually transmitted infections (STIs) are a major health concern globally. It is estimated that in the US $16 billion is spent annually on STI related diseases, with direct costs in treating infections totalling to $740 million.[163] In 2012, the estimated worldwide prevalence of *C. trachomatis* was 4.2% in women and 2.7% in men,[164] with other common STIs (*e.g*. *N. gonorrhoea,* trichomoniasis, syphilis) typically having 0.5–1% prevalence,[165] while there are roughly 1 million new curable STI infections every day.[166] Patients with STIs have increased susceptibility to HIV, due to associated inflammation or ulcers.[167]

Like much of the world, diagnosis rates of STIs in England has risen over the last decade, (20% increase in new STI diagnoses)[168] however this cannot be wholly attributed to increased transmission rates. Increased awareness towards testing amongst all age groups has resulted in an increase of access to sexual health services, with total visits to specialist health services increasing by 386,000 (15.2%) between 2012 and 2016.[169]

Efficient testing protocols are essential to enable timely and effective treatment. As discussed in Chapter 1, a range of testing protocols have been developed, such that the diagnosis of most pathogens can be achieved in only a few days. Regardless of the method in which the pathogen is detected (culture, immunoassay, NAAT, *etc.*), tests are typically performed in central laboratory settings, and as such have a significant time delay from the initial appointment to results and the administration of antibiotics (Scheme 4.1). Rapid point-of-care tests (POCTs) allow for decentralised diagnosis and can therefore offer significantly reduced waiting times for patients.

The speed of diagnosis can have significant impact on the patient's view of treatment, with some reports suggesting 99% of patients are willing to wait 2 hours for *C. trachomatis* results,[170] and 98.8% would wait 1 hour for *N. gonorrhoeae* results.[171] There is some discrepancy, however, with other reports indicating only 50% of patients would be willing to wait 40 minutes for CT diagnosis.[172] Despite the differences in statistics, it is clear there is a market for rapid POCTs.



Scheme 4.1 Typical diagnosis pathway (adapted from Atlas Genetics marketing material)

Nucleic acid detection methods offer improvements over culture methods not only because of their speed, but the increased information that can be obtained could offer quicker treatment for critical, time-sensitive conditions such as sepsis or bacterial meningitis. The knowledge of specific serotypes can improve treatment protocols, allowing personalised medicine. For example, *N. gonorrhoeae* (NG), amongst other bacteria, exhibits many different strains globally, each with varying antibiotic resistance. Infections caused by *Staphylococcus* bacteria are one of the most common and most can be treated with common antibiotics, however *S. aureus* has developed resistance to many antibiotics and is notoriously difficult to treat.[173] Diagnosis of the bacteria's serotype, and determination of which antibiotic to use, is usually conducted *via* culture followed by *in vitro* susceptibility testing, however rapid NAAT tests would eliminate the requirement for susceptibility testing, given the site of mutagenesis is known.[16]

4.1.1 Point-of-Care diagnostics

POCTs allow users with limited skills to execute highly complex analysis with minimal training. Typically, the healthcare worker which has the face-to-face interaction would not carry out any analysis as they do not have the required equipment, training, or time. Through the development of user-friendly devices which require little to no skilled input, centralised testing can be made more accessible with reduced waiting times.

A number of rapid POC tests are now commercially available, however few have been widely implemented in their target market and routine testing can still take up to 9 days to complete.[174] Early evaluations of implemented POC tests in England show a potential £11.7 million saving per year to the NHS,[175] as well as preventing up to 17,000 new infections and nearly 100,000 overtreatments compared to standard care.[175] Several commercial POCTs

based on nucleic acid amplification exist for pathogenic bacteria,[15] while a range of non-NAATs exist for both CT,[176] and NG[177] with varying accuracy.

The challenges facing POC devices varies between the developed and the developing world. Limited-resource settings use benchmark ASSURED criteria to identify the most appropriate testing equipment.[178] This requires devices to be Affordable, Sensitive, Specific, User-friendly, Rapid and robust, Equipment-free and Deliverable to the end-users.[12] Of course, sacrifices must be made for different criteria, but ultimately the devices must be reliable and available to users where other resources are limited. Devices are often limited due to their power needs or portability, particularly in rural areas, whereas developed countries have fewer constraints in the requirements of equipment, such that benchtop, rather than handheld, devices are usually sufficiently small and portable.

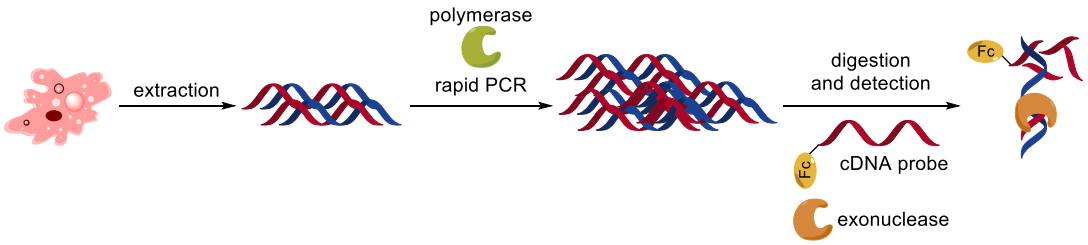
Although cost reductions can be achieved through implementation of rapid POCTs, implications of false-negatives and false-positives remain a challenge. False-positives result in increased administration of unnecessary antibiotics, potentially contributing to the ongoing antibiotic crisis.[173] False-negatives result in increased susceptibility to other infections and complications caused by the bacteria, as well as further transmission of infections to other patients. False-negatives, in the long term, can be costly due to complications requiring additional treatment.

Sensitivity and specificity of POC tests is typically lower than gold standard assays,[179] however the improvement on the speed of patient treatment can outweigh the need for high sensitivity and specificity and result in an increase in the number of patients that receive treatment.[180]

4.1.2 Atlas Genetics' Diagnostic Assay

Commercial POC NAATs mainly use fluorescent detection methods,[174] however no device has been widely implemented in healthcare settings. Atlas Genetics' (AG) proprietary technology combines chemistry, biology, and engineering into a single device to enable efficient patient diagnosis in a reduced time frame. AG have a range of patents covering their technology, ranging from the microfluidic system to the chemical labelling of the oligonucleotides.[116,117,155,181–184]

The diagnostic assay is based on PCR amplification and solution-based electrochemical detection (Scheme 4.2), combining the sensitivity and specificity of PCR with the affordability of solution-based electrochemistry.



Scheme 4.2 Atlas Genetics' diagnostic assay method

The test requires minimal user intervention. The healthcare worker takes a clinical sample then loads the sample directly onto the disposable cartridge. This is then loaded into the desktop machine, the io®, where the patient's details are input (Figure 4.1). From this point, the test is autonomous, with results displayed on screen showing either "positive" or "negative" for the desired target.

Figure 4.1 Atlas Genetics io and cartridge

The detection method allows for detection of essentially any DNA, given the target sequence is known. This lends itself well to applications for not only STIs, but also hospital acquired infections and genetic biomarkers. A range of tests are under development, with the CT assay achieving CE Marking in 2015, while the combined CT/NG assay is set to achieve CE Marking in 2019 and FDA approval in 2020. The CT assay offers sensitivity of 98.1% and specificity of 98.0%,[185] while pre-clinical trials for the combined CT/NG assay showed a sensitivity of 91.9% and a specificity of 98.3% for CT, and sensitivity of 94.1% and specificity of 99.8% for NG.

The development of a library of labels has enabled a 6-plex detection system, exploiting the adjustable oxidation potential of ferrocene, however only two of these labels have been employed for use in a diagnostic assay due to limited sensitivity of the majority of the library. As such, the CT/NG assay utilises multiple detection chambers to allow for simultaneous detection of several targets, detecting for the internal control and a single target DNA sequence in each chamber.

Improved digestion performance of probes could enable the use of multiple labels in a single chamber which have been previously unsuitable, therefore improving the multiplex capability of the system. If the improvements observed in Chapter 2 and Chapter 3 were also displayed in a full DNA detection assay, the detection of many more targets could be realised in a single test. In addition to improved multiplex capability, increased probe sensitivity could allow for fewer PCR cycles to be conducted, thus decreasing the time required to run the assay.

4.2 Double stranded digestions of internally labelled probes

There is clear need for POC diagnostics in both the developed and developing world, with rapid diagnostics offering increased public health and reduced cost to healthcare providers. Nucleic acid amplification tests are superior to other methods (*e.g.* antigen detection) due to their sensitivity. Combining NAAT technology with solution-based DNA detection offers an opportunity to create sensitive, reliable, and cheap tests which could be widely implemented.

It was of interest to investigate whether the introduction of internal labels into NAAT based tests could offer improved detection sensitivity, thus improving the performance of the diagnostic assays. The proof-of-concept single stranded assays discussed in Chapter 1 and Chapter 2 displayed promising results, suggesting that the internal labelling strategy could result in improved performance in double-stranded assays.

After successful single stranded digestions, the probes were substituted into double stranded assays for future incorporation into DNA detections. There are a number of stages where the modified probes could fail in the assay. Firstly, the probes must hybridise with the complimentary DNA, as T7 exonuclease (T7 or T7 exo) will only digest dsDNA.[186] Secondly, the probes must not inhibit T7's activity and allow the release of smaller units into solution.

For reduced complexity, initial studies were conducted on synthetic target (ST) DNA, which are complementary oligonucleotides which have been made on a DNA synthesiser, rather than enzymatically (*e.g. via* PCR). This allows the probes to be studied with increased complexity compared to the ssDNA assays, but does not introduce such complex matrices associated with using PCR to prepare the complimentary DNA strands.

Assays are conducted by combining a "detection mix" containing the appropriate probe and T7 exonuclease in T7 buffer with the target DNA. These are then heated to 37 °C for 3 minutes using either thermal cyclers and an Autolab potentiostat for detection, or digesting the probes directly on the io machine. The digestion protocol was not optimised for internal probes, instead directly adopting the digestion protocol developed by AG for use in their assays, to see how these probes perform in comparison to the AG in-house standard.

4.2.1 Internal Control double stranded assays

The modified derivatives of AG's internal control probes were first studied, using a "no template control" (NTC) to study the background in the new buffer, and to assess the reactivity between the internally labelled ssDNA probe and T7 exo (Figure 4.2). As the synthetic target negatives are completely DNA free and T7 is inactive on ssDNA (to any

reasonable degree), the negatives look to determine the inherent current that can be generated from probes in the T7 buffer.

Probe	Sequence
IC-C	5'Y-GCA CGA TCC CTT TCC TAA AGA CG-3'
IC-1	5'-[p]-GCA CGA U^{m144}CC CTT TCC TAA AGA CG-[p]-3'
IC-2	5'-[p]-GCA CGA U^{m144}CC CU^{m144}T TCC TAA AGA CG-[p]-3'
IC-3	5'-[p]-GCA CGA U^{m144}CC CU^{m144}U^{m144} TCC TAA AGA CG-[p]-3'
IC-4	5'-[p]-GCA CGA U^{m144}CC CU^{m144}U^{m144} U^{m144}CC TAA AGA CG-[p]-3'
IC-5	5'-[p]-GCA CGA U^{m144}CC CU^{m144}U^{m144} U^{m144}CC U^{m144}AA AGA CG-[p]-3'

Table 4.1 Internal Control oligonucleotide sequences

In initial testing, a probe concentration of 6 μM was utilised. This was too low to give a detectable signal, and as such the probes were studied at the increased concentration of 12 μM. The background current in the T7 buffer was significantly lower than that of the S1 buffer, despite doubling the concentration of the probes. The background of IC-C, IC-1 and IC-2 were either too low to detect, or under 5 nA even at 12 μM. (*cf.* in the range of 10^2 nA for S1 nuclease). Concurrent with S1 buffer, the signal of IC-3, IC-4, and IC-5 increased in a linear fashion such that IC-5 had the greatest background signal. The background signal is believed to be supressed slightly by the addition of bovine serum albumin (BSA) and Triton added to the detection mix.

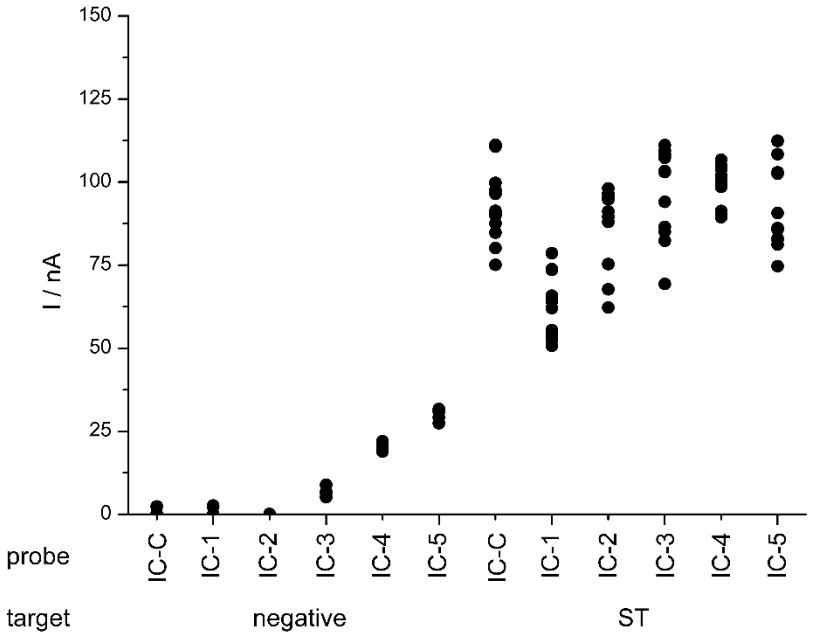


Figure 4.2 Internal Control probes, double stranded digestions. Detection mix: Probe (12 μM), detection buffer, T7 exo (16 U). DNA mix: NTC ST mix (20 μL) or IC ST (12 μM) mix (20 μL). 37 °C, 3 minutes in tube, Autolab detection.

The digestion of the probes in T7 buffer was then studied. The probes were treated with synthetic target (12 μM) and heated in PCR tubes at 37 °C for 3 minutes followed by electrochemical analysis at room temperature. Crucially, and most importantly, the probes gave a positive result in the presence of both T7 exo and complementary DNA, showing that the probes undergo hybridisation and are susceptible to enzymatic digestion.

It was, however, disappointing to observe no significant difference in signal between any of the probes, or if anything slightly worse performance when comparing IC-C and IC-1. This was in contrast to the results observed for the S1 digestions, where increasing the level of substitution led to a concurrent increase in the current generated. As T7 and S1 have different directions of action, with T7 working in the 5'-3' direction, this could account for the difference in signal observed between these two digestion protocols. Additionally, the different buffer solutions could account for the reduction.

Synthetic target digestions give a crude indication of whether the dsDNA assay will be successful, as it is a simplified version of the full diagnostic assay. As such, to investigate the properties of the new probes in a full assay mimic, DNA was amplified *via* PCR to increase the complexity of the system, instead of using purified synthetic target DNA. To this extent, 10 pg IC DNA was amplified *via* the rapid PCR protocol developed by AG. The probes were then studied under identical conditions to the ST assays, but replacing the synthetic target mix for PCR amplicon.

Again, the background was first studied, and there was some degree of signal suppression observed compared to the ST assays. The additional complexity of the PCR solution (primers, polymerases *etc.*) is possibly limiting the availability of the ferrocene at the electrode surface, acting as blocking agents. The trend of the background oxidations, however, remained the same in the PCR amplicon assay compared to the synthetic target.

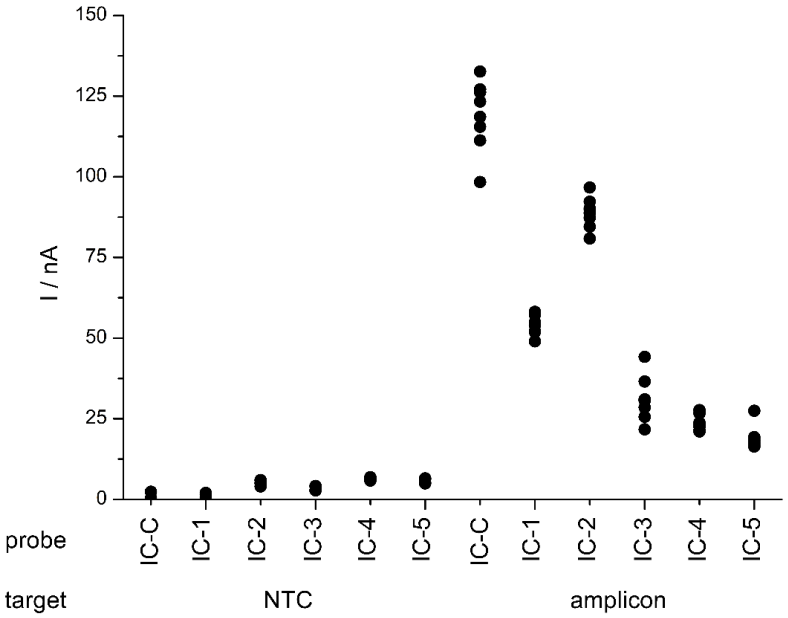


Figure 4.3 Internal Control probes, double stranded digestions. Detection mix: Probe (12 μM), detection buffer, T7 exo (16 U). DNA mix: NTC PCR mix (20 μL) or IC PCR (10 pg) mix (20 μL). 37 °C, 3 minutes in tube, Autolab detection.

The IC probes were then treated with IC amplicon produced *via* PCR, which were then incubated in PCR tubes at 37 °C for 3 minutes followed by detection *via* DPV at room temperature (Figure 4.3). While all probes still gave positive results in the presence of cDNA, the results were somewhat disappointing compared to the synthetic target digestions, resulting in a lower signal than the control IC-C for all probes.

These digestions showed that these probes were unable to give any increase in signal compared to traditional 5' labelling. In efforts to increase the poor signal observed during the T7 digestions, the probes were next analysed using AG's custom-build potentiostat, which is coupled to a Peltier in order to conduct digestions directly on the electrode sub-circuit, and also allows the detections to be conducted at elevated temperatures.

As such, the digestions were repeated, instead conducting both the digestion and detection on the sub-circuit surface. The samples are held at 37 °C for 3 minutes to allow for T7 digestion, followed by DPV analysis while holding at 37 °C.

The main downfall of the custom-build potentiostat is that both the hardware and software are inferior to that of the Autolab, and as such the DPVs generated are not of such good quality. The increased scan rate results in signals which have significantly more noise. Additionally, the code which was written for the peak analysis is very sensitive to peak width and the potential around which it is centred, meaning the algorithms written for AG are not optimised for use with the internally labelled oligonucleotides developed in this

thesis. This results in a slightly lower signal than would be expected by visual inspection, as part of the peak is included in in the calculation for the baseline. The move to the io also resulted in apparent elimination in background current, as this can often no longer be distinguished from the noise of the instrument by the algorithm used to calculate peaks.

Upon digestion, there was an increase in signal observed for all probes due to the elevated detection temperature (Figure 4.4). Probe IC-2 containing two substitutions offered a slight increase in signal over the 5'-labellec control IC-C, however the digestion of the other probes containing multiple substitutions could not improve the signal generated.

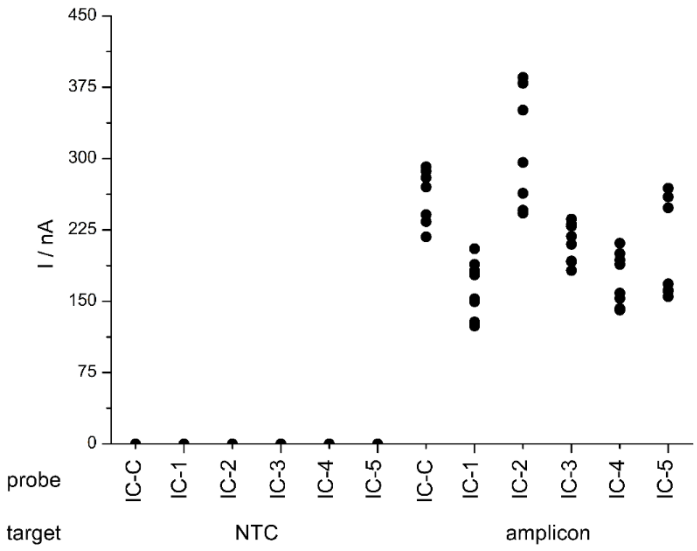


Figure 4.4 Internal Control probes, double stranded digestions. Probe (12 μM), detection buffer, T7 (16 U), IC PCR (10 pg) mix (20 μL). 37 °C, 3 minutes, io detection.

The IC probes were the first internally labelled probes to be subjected to dsDNA digestions. While it was disappointing that the new probes did not yield an increase in signal compared to the control probe, it was promising to observe digestion for all probes under all conditions tested. The experiments in this chapter showed that crucially, the internal labels did not inhibit duplex formation or T7 exonuclease action.

4.2.2 *Chlamydia trachomatis* double stranded assays

As previously discussed in Chapter 2, the sequence selection of *C. trachomatis* was chosen to ensure as much space between labelled nucleosides as possible. Although S1 digestions of the mono-ferrocenyl CT probes showed a reduction in current compared to the IC sequence, an increase in signal of the assay was still a possibility due to the different mode of action between enzymes T7 *vs.* S1.

4.2.2.1 Mono-ferrocene *C. trachomatis* probes

To this extent, the mono-labelled CT probes were first subjected to T7 digestion with synthetic target DNA. The backgrounds of the probes were studied at the increased concentration of 12 μM in T7 buffer, and were of a similar level in current to the IC background signal. The CT probes CT-1 to CT-5 displayed increasing current in a linear fashion such that CT-5 gave the highest background signal (Figure 4.5).

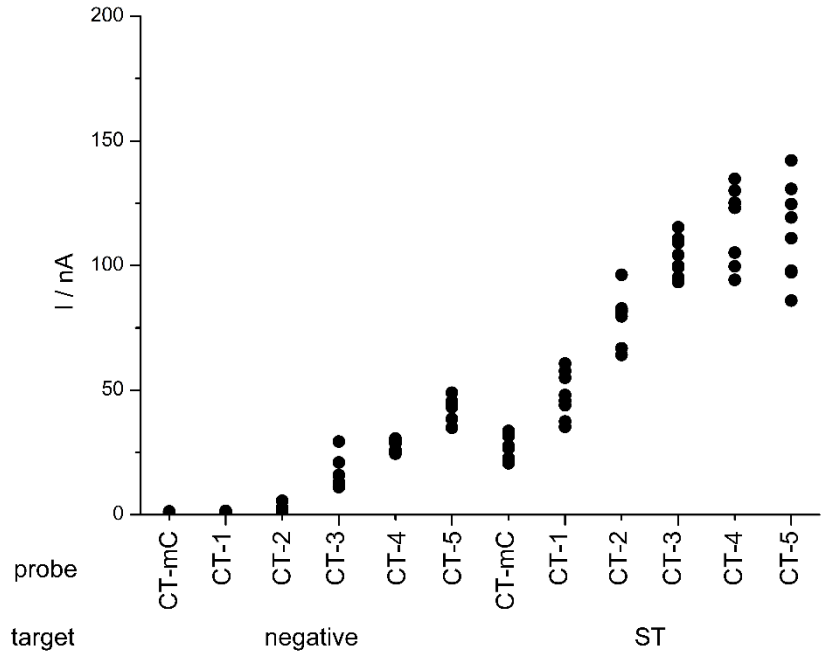


Figure 4.5 mono-CT probes, double stranded digestions. Detection mix: Probe (12 μM), detection buffer, T7 exo (16 U). DNA mix: NTC ST mix (20 μL) or CT ST (12 μM) mix (20 μL). 37 °C, 3 minutes in tube, Autolab detection

Probe	Sequence
CT-mC	5'Y - CTG TCC GCT GGT TCT TCC TTA CT-[p]-3'
CT-1	5'-[p]-CU^{m144}G TCC GCT GGT TCT TCC TTA CT-[p]-3'
CT-2	5'-[p]-CU^{m144}G TCC GCU^{m144} GGT TCT TCC TTA CT-[p]-3'
CT-3	5'-[p]-CU^{m144}G TCC GCU^{m144} GGT TCT TCC TTA CU^{m}-[p]-3'
CT-4	5'-[p]-CU^{m144}G U^{m144}CC GCU^{m144} GGT TCT TCC TTA CU^{m}-[p]-3'
CT-5	5'-[p]-CU^{m144}G U^{m144}CC GCU^{m144} GGT TCT TCC U^{m144}TA CU^{m144}-[p]-3'

Table 4.2 mono-Ferrocene *C. trachomatis* oligonucleotide sequences

The probes were then treated with T7 exonuclease in the presence of CT synthetic target, digesting in PCR tubes at 37 °C for 3 minutes. Pleasingly, the digestion of all internally labelled probes gave increased signal compared to the 5'- labelled control probe, CT-mC, increasing in current in a manner similar to the S1 digestions. CT-1 gave increased signal,

followed by further increases in signal for all probes, however there was a plateau with probes CT-4 and CT-5, showing no significant increase compared to CT-3.

These results were promising, showing that with synthetic target DNA, the signal of the system can be increased upon digestion due to an increased number of labels. This data set also gave some evidence towards the hypothesis that the labelled nucleosides should be spread out to give optimum signal, as the IC sequence has three adjacent labelled nucleotides, which is thought to inhibit the assay.

After these promising results, the CT probes were subjected to digestion with PCR amplicons of *C. trachomatis* serovar F (100 IFUs). As observed with the IC probes, there was a slight suppression in background current when using PCR mix, rather than synthetic target (Figure 4.6). Again, this could be attributed to a possibly blocking effect of the additional components on the availability of the probe at the electrode surface.

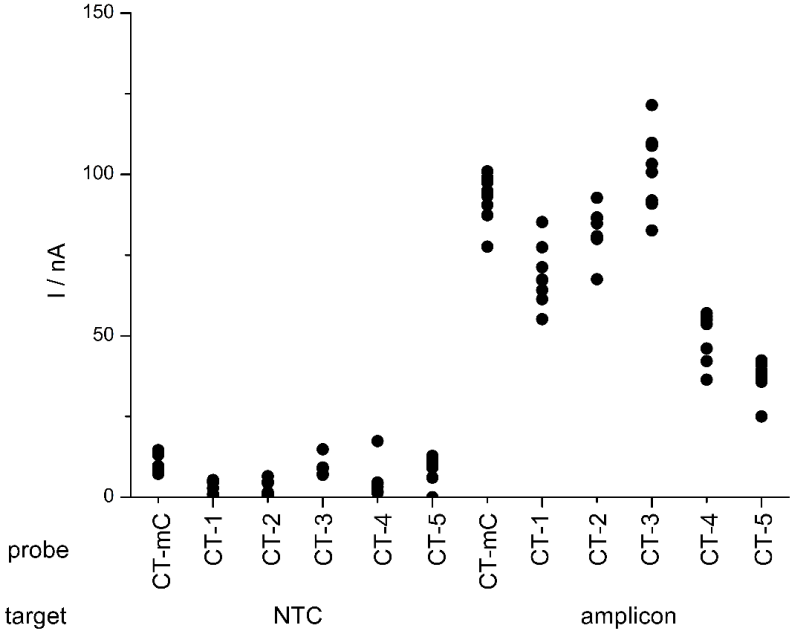


Figure 4.6 mono-CT probes, double stranded digestions. Detection mix: Probe (12 μM), detection buffer, T7 exo (16 U). DNA mix: NTC PCR mix (20 μL) or CT PCR (100 IFUs) mix (20 μL). 37 °C, 3 minutes in tube, Autolab detection.

The amplicon solution was then treated with detection mixture and digested in PCR tubes at 37 °C for 3 minutes (Figure 4.6). Compared to IC, the CT probes gave higher current upon digestion, although this was still significantly lower than signals that would be generated in a standard assay. Digestions for the control showed higher signals than CT-1 and CT-2, with CT-3 giving similar signals to the control. It was clear, however that there was a significant reduction in signal for CT-4 and CT-5 which was not observed in the synthetic target studies. In contrast to synthetic target digestions the internal probes gave no significant increase in

signal in comparison to the control probe, but still exhibited activity towards T7 exonuclease upon introduction of cDNA.

Further CT dsDNA digestions were carried out on the io sub-circuits to mimic the commercial assay digestion protocol, with digestion and detection occurring directly on the electrodes. In a similar manner to the IC, when detections were carried out on the io, it resulted in reduced background current and an increase in signal of around 100 nA due to the increased detection temperature. Again, the internal probes showed a slight reduction for CT-1 and CT-2, however the digestion of probes CT-4 and CT-5 gave significantly improved results, generating similar signals to the control probe (Figure 4.7).

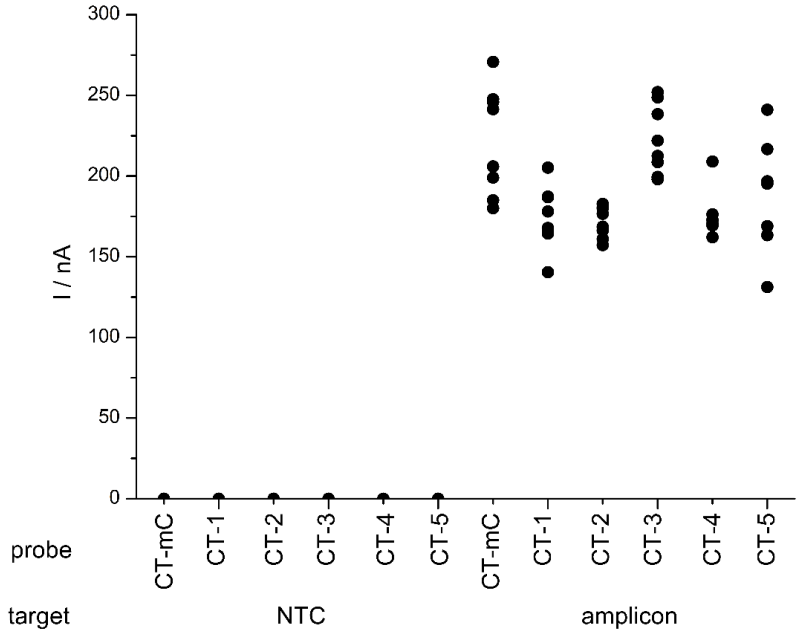


Figure 4.7 mono-CT probes, double stranded digestions. Probe (12 μM), detection buffer, T7 (16 U), CT PCR (100 IFUs) mix (20 μL). 37 °C, 3 minutes, io detection.

The increase in signal observed from the move from Autolab detection to io detection was larger than expected, but was displayed across both IC and mono-CT probes. The discrepancy in currents generated upon digestion between the Autolab and io is attributed to the increased temperature, however additional benefit could be due to the proximity of probes to the electrode prior to detection. The very low levels of potential across the electrode could allow for pre-equilibration during digestion, therefore leading to higher concentration at the electrode surface.

The mono-CT probes showed promising results across all dsDNA digestions, showing improved performance over the IC probes. This gave some evidence towards the optimal labelling pattern. IC has labelled uridine units adjacent to each other, which is believed to

cause some level of disruption in base-pairing ability and T7 activity. The labelling pattern of the mono-CT probes, however, is much more spread out, with no labels being adjacent to each other. This was believed to improve the performance of the probes upon treatment with T7 exonuclease.

4.2.2.2 Di-ferrocene *C. trachomatis* probes

The mono-ferrocene probes showed improved digestion performance over the IC sequence which was attributed to the spacing of the labelled nucleotides, however the concentration of probe required was still relatively high compared to AG typical probe concentration in an assay (12 μM *vs.* 6 μM in a typical assay). Due to the di-ferrocene probes' increased sensitivity over the mono-ferrocenes in S1 digestions discussed in Chapter 2 and Chapter 3, it was investigated whether the di-ferrocene probes could offer increased signal in dsDNA assays. Due to the increased sensitivity exhibited by these probes in an S1 digestion compared to the mono-labels, the concentration of the probe was reduced to 6 μM, to mirror the typical probe concentration in an AG diagnostic assay.

To mimic the diagnostic assay, the di-ferrocene CT probes were treated with PCR amplicon to investigate whether the di-ferrocene CT internally labelled probes could be used to increase the signal in dsDNA assays compared to the control probe.

To this extent, the background of the probes was studied in NTC PCR mix, and concurrent with previous studies the probes gave an increase in background current for each sequential dT substitution (Figure 4.8). The background currents generated were a similar level to the mono-ferrocene CT equivalents despite the reduced concentration of the di-Fc CT probes.

Probe	Sequence
CT-dC	5'-**Z**- CTG TCC GCT GGT TCT TCC TTA CT-[p]-3'
CT-6	5'-[p]-C$\underline{U^{d275}}$G TCC GCT GGT TCT TCC TTA CT-[p]-3'
CT-7	5'-[p]-C$\underline{U^{d275}}$G TCC GC$\underline{U^{275}}$ GGT TCT TCC TTA CT-[p]-3'
CT-8	5'-[p]-C$\underline{U^{d275}}$G TCC GC$\underline{U^{275}}$ GGT TCT TCC TTA C$\underline{U^{d275}}$-[p]-3'
CT-9	5'-[p]-C$\underline{U^{d275}}$G $\underline{U^{d275}}$CC GC$\underline{U^{d275}}$ GGT TCT TCC TTA C$\underline{U^{d275}}$-[p]-3'
CT-10	5'-[p]-C$\underline{U^{d275}}$G $\underline{U^{d275}}$CC GC$\underline{U^{d275}}$ GGT TCT TCC $\underline{U^{d275}}$TA C$\underline{U^{d275}}$-[p]-3'

Table 4.3 *C. trachomatis* probe di-ferrocene sequences

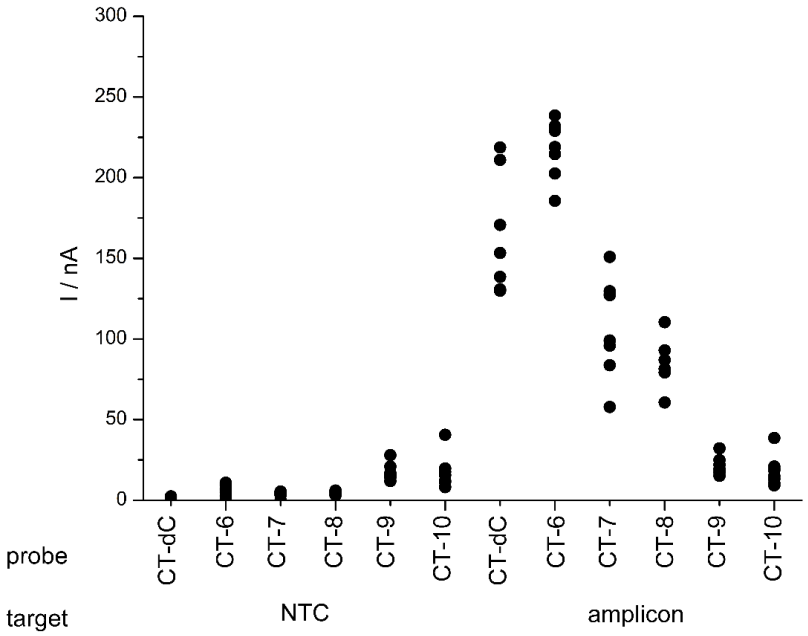


Figure 4.8 di-CT probes double stranded digestions. Detection mix: Probe (6 μM), detection buffer, T7 exo (16 U). DNA mix: NTC PCR mix (20 μL) or CT PCR (100 IFUs) mix (20 μL). 37 °C, 3 minutes in tube, Autolab detection.

CT PCR amplicon mixture was then treated with detection mixture and heated to 37 °C for 3 minutes and analysed *via* DPV on the Autolab (Figure 4.8). The probes displayed increased signal *vs.* their mono-ferrocene equivalents, allowing detection to be carried out at the standard probe concentration of 6 μM while still offering an increase in signal. Probe CT-6, containing a single U^{d275} substitution, offered a slight increase in signal compared to the control probe, CT-dC, however further substitution of the probe resulted in significant reduction in signal. The probes containing four and five substitutions (CT-9 and CT-10) failed to digest in the given conditions, displaying neither a shift in E_{ox} or increase in current.

The probes were then incorporated into io digestion and detection to mimic the digestion protocol used in an AG diagnostic assay (Figure 4.9). An increase in signal was again observed from moving from the Autolab to the io potentiostat despite the same concentration of probe used. This increased the current of the probes containing two and three substitutions such that CT-dC, CT-6 and CT-7 showed a similar spread of results, with a slight reduction in signal for CT-8, however the probes containing four and five U^{d275} units remained inactive to T7 exonuclease. Due to their inactivity, CT-9 and CT-10 were omitted from any further studies due to inactivity. The complete inactivity suggests these probes are unable to form a duplex and are therefore left intact by T7.

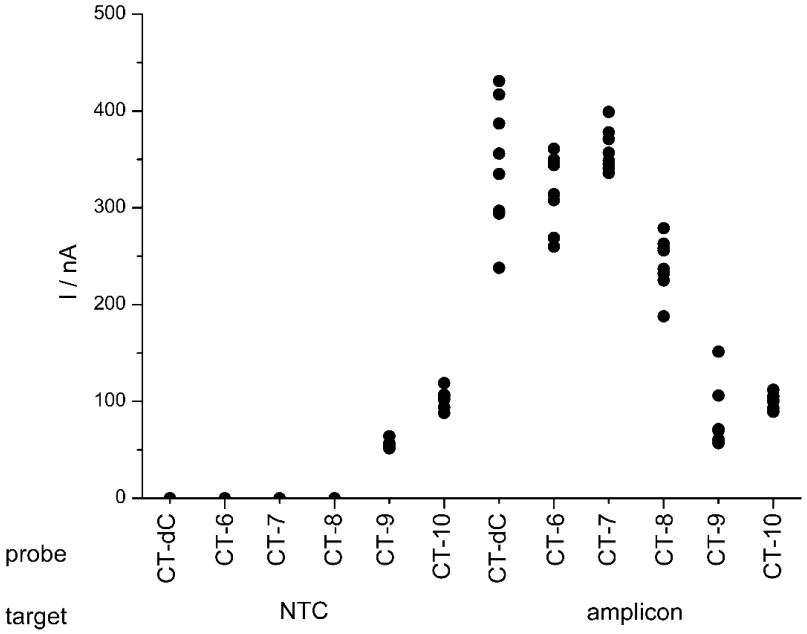


Figure 4.9 di-CT probes, double stranded digestions. Detection mix: Probe (6 µM), detection buffer, T7 exo (16 U). DNA mix: NTC PCR mix (20 µL) or CT PCR (100 IFUs) mix (20 µL). 37 °C, 3 minutes, io detection.

In a clinical assay, all the reagents would be pre-loaded into a disposable cartridge. To mimic this, the detection mixture containing the probe was applied to Atlas io sub-circuits and dried onto the surface prior to detection. To achieve this, a detection mix spot of 4 µL was applied to the centre of the sub-circuit and dried at 45 °C, 10% relative humidity for 10 minutes. The sub-circuit was then sealed with adhesive and the screen-printed electrodes (Figure 4.10).

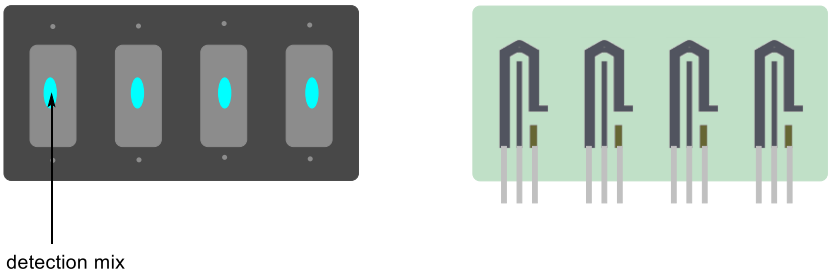


Figure 4.10 Atlas Genetics sub-circuit and screen-printed electrodes

These "dried-down" cartridges are reconstituted prior to use by applying the PCR amplicon mixture into the opening and mixing between the rest position and first stop of the pipette three times (Figure 4.11). This replicates the microfluidic movement of solutions through the cartridge in a full assay. The sub-circuits are then placed in the heated detection rig and subjected to the same digestion protocol as before, such that the probes are heated at 37 °C for 3 minutes for digestion followed by simultaneous detection *via* DPVs at 37 °C across 4 channels.

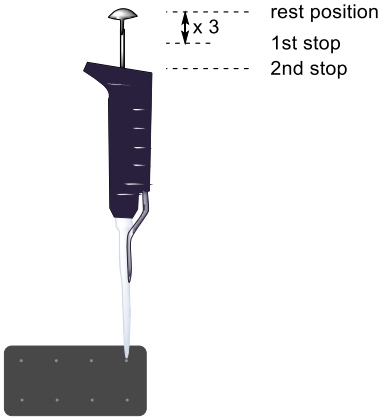


Figure 4.11 Loading of sample to dried-down sub-circuits

The move from wet to dry detection mixture resulted in a large increase in signal of around 200 nA generated for the probes analysed (Figure 4.12). The control, CT-6, and CT-7 gave nearly identical digestion signals, while CT-8 resulted in slightly reduced signal, as observed with the previous io detection (*cf.* Figure 4.9). Although there was an improved signal using the dry detection mix, there was no benefit to using the internally labelled system compared to the traditional 5'-labelling.

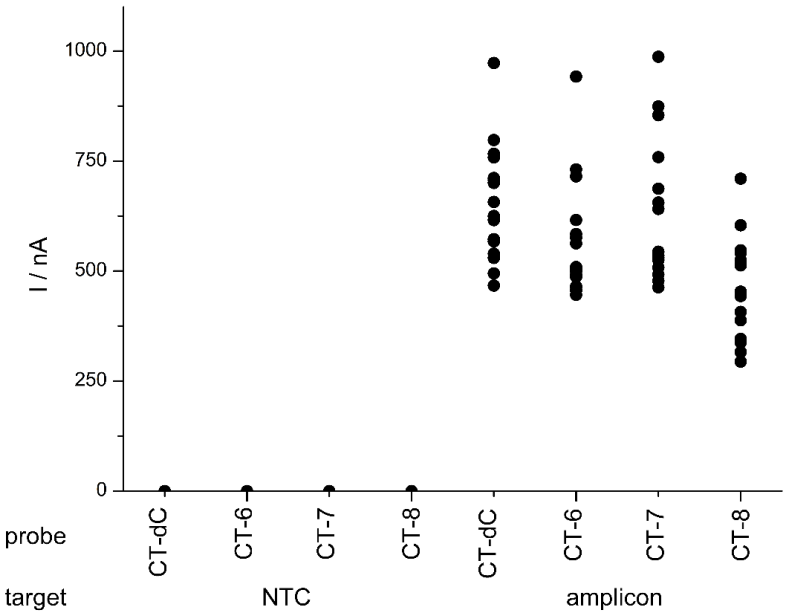


Figure 4.12 di-CT probes, double stranded dried detection. Detection mix (dry down, sub-circuit): Probe (6 µM), detection buffer, T7 exo (16 U). DNA mix: NTC PCR mix (20 µL) or CT PCR (100 IFUs) mix (20 µL). 37 °C, 3 minutes, io detection.

Due to the similarity of double stranded digestion results observed across all probes, it was thought that the mechanism of T7 exonuclease digestion may result in incomplete digestion

of the probe strand, therefore limiting the signal that could be generated despite the increased ferrocene substitution.

It is hypothesised that while the first label is being released, the next labels further down the sequence are not being digested, instead remaining as long oligonucleotides. Upon cleavage at the 5' end, and therefore shortening of the probe, the duplex will be destabilised due to a reduction T_m, and as such the probe will be displaced and T7 will no longer be active on the single strand.

The probes all showed promising results that it is indeed possible to digest the internally labelled probes, however these could not be used to increase signal in a full AG style assay using T7 exonuclease.

4.2.3 Labelling patterns and mechanistic insight into T7 exonuclease digestion

The exact mechanism of T7 exonuclease is currently unclear, and as such it is unclear what digestion products are being released after digestion (*i.e.* mono, di, tri-nucleotides). If this could be determined, it would allow more informed decisions to be made on probe sequences and allow for intelligent design of probes.

The first CT sequences were designed with the hypothesis that ensuring the labels were as spread out as possible would enable digestion. It was decided to investigate the effect of grouping the substitutions closer together and determining whether the probe would still digest. The original probes labelled base positions 2 and 9, and 2, 9 and 23, so two new probes were designed to condense the spacing between the labels, and labelling bp positions 2 and 4, and 2, 4 and 9. This could give insight into optimal label sequencing.

Probe	Sequence
CT-7	5'-[p]-C$\underline{U^{d275}}$G TCC GC$\underline{U^{275}}$ GGT TCT TCC TTA CT-[p]-3'
CT-8	5'-[p]-C$\underline{U^{d275}}$G TCC GC$\underline{U^{275}}$ GGT TCT TCC TTA C$\underline{U^{d275}}$-[p]-3'
CT-11	5'-[p]-C$\underline{U^{d275}}$G $\underline{U^{d275}}$CC GCT GGT TCT TCC TTA CT-[p]-3'
CT-12	5'-[p]-C$\underline{U^{d275}}$G $\underline{U^{d275}}$CC GC$\underline{U^{d275}}$ GGT TCT TCC TTA CT-[p]-3'

Table 4.4 *Chlamydia trachomatis* probe sequences

These probes were digested *via* S1 digestion to investigate the theory that the closer the labels are together, the larger the digested products and therefore lower the signal. Both new probes CT-11 and CT-12 were successfully digested by S1 nuclease while displaying reduced current compared to the labels which were more spread out (*cf.* CT-7 and CT-8) (Figure 4.13). This supports the hypothesis that the probes are required to be more spread out to enable efficient digestion.

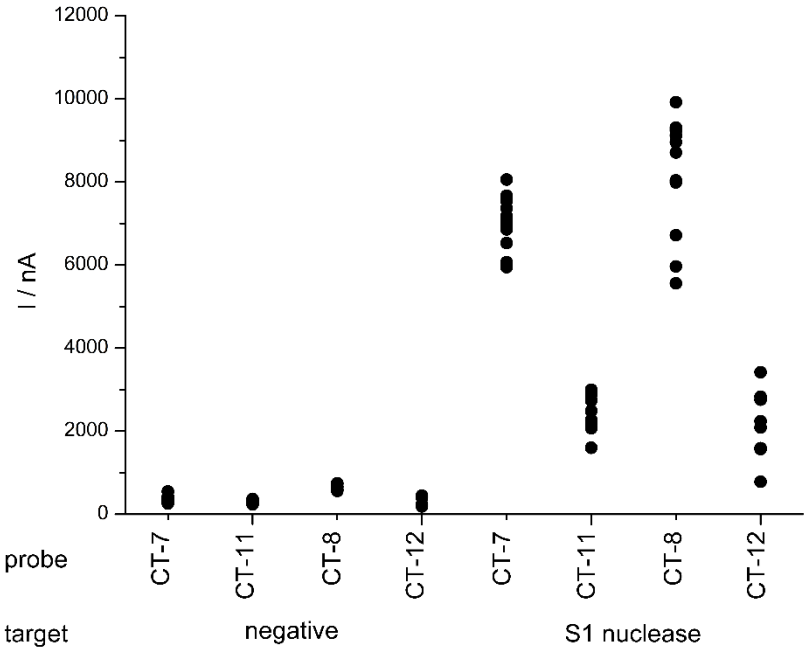


Figure 4.13 di-CT probes, S1 digestions. Probe (3 μM), pH 4.5 NaOAc buffer. Negative control or S1 (10 U). 37 °C, 30 minutes in tube, Autolab detection.

These probes were then subjected to double stranded digestion with PCR amplicon under standard conditions, however neither CT-11 nor CT-12 showed any increase in current for amplicon *vs.* NTC, suggesting the probes were unable to form a duplex due to their inactivity.

An additional di-ferrocene CT probe containing a single U^{d275} substitution, CT-13, was designed such that the U^{d275} unit was at the 3' terminus on the 23rd base to investigate the effect of terminal labelling. It was hypothesised that by introducing the label at the terminus, T7 exonuclease would not fully digest this probe and no signal would be generated, as the duplex would be insufficiently stable due to its shortened length.

Probe	Sequence
CT-13	5'-[p]-CUG UCC GCU GGT TCT TCC TTA C$\underline{U^{d275}}$-[p]-3'

Table 4.5 3'-terminus labelled CT probe

Treatment of CT-13 with S1 nuclease resulted in a reduction in current compared to CT-6 (Figure 4.14). The background oxidation was, as expected, of a similar level to that of CT-6, however upon digestion the probe did not result in a significant increase in current generated, being a factor of ten times lower than the CT-6 digest. Due to the proximity of the labelled nucleotide to the 3'-end, this is possibilty inhibiting the 3'-5'- action of S1 nuclease.

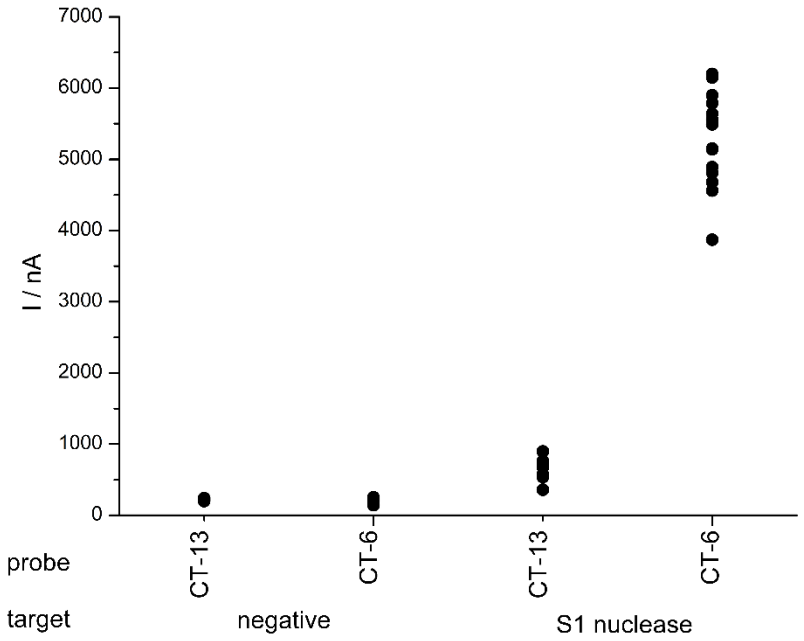


Figure 4.14 di-CT, probes S1 digestions. Probe (3 μM), pH 4.5 NaOAc buffer. Negative control or S1 (10 U). 37 °C, 30 minutes in tube, Autolab detection.

The probe was then treated with CT PCR amplicon however it was shown to be completely inactive with no change in E_{ox} or current observed. This is supports the hypothesis the formation of a shorter sequence will reduce duplex stability, therefore denaturing at a lower temperature and being unable to be digested as far down as the label.

These studies have shown that for successful duplex formation and digestion, the probes must have more than one unsubstituted base in between them. If the labelled base is too close to the terminus, the probe will not undergo complete digestion as the T7 does not fully digest the labelled oligonucleotide.

4.3 Multiplex detections with internally labelled oligonucleotides

Double stranded digestions of the internally labelled probes with T7 exonuclease did not result in an improvement over traditional oligonucleotide labelling strategies. It was sought to determine whether this strategy could be used to increase the sensitivity of existing AG labels through improved digestion properties. In AG's multiplex system, they have previously utilised two di-ferrocene labels for duplex detection. Other labels designed have lacked sensitivity therefore generating weak signals upon digestion and as such have not been successfully implemented into a triplex detection. For further development, it is desirable to create a library of labels with similar sensitivity to enable accurate diagnosis.

The di-ferrocene deoxyuridine units previously synthesised allow for three labels to be detected in a single voltammogram. It was sought to create a multiplex detection system with the internally labelled ferrocenyl probes, and to this extent probes were designed to detect for IC, CT, and NG targets simultaneously. AG use two sequences for the detection of *N. gonorrhoeae*, however one was deemed to be less suitable for testing due to the location of the thymidine residues.[187]

Probe	Sequence
IC-098	5'-[p]- CGG U^{d098}GT AGT TAT CTT GAC TCC TCC CGA GC -[p]-3'
NG-437	5'-[p]- GCA CGA U^{d437}CC CTT TCC TAA AGA CG -[p]-3'
CT-6	5'-[p]-CU^{d275}G TCC GCT GGT TCT TCC TTA CT-[p]-3'

Table 4.6 Multiplex probes for internal control, *C. trachomatis*, and *N. gonorrhoeae*

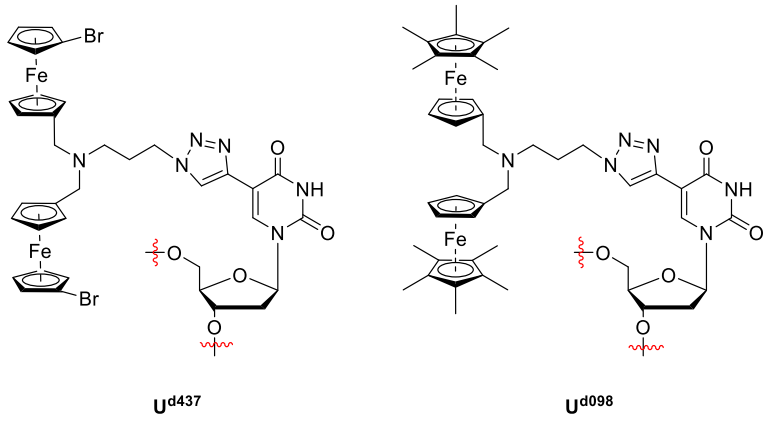


Figure 4.15 Key for Table 4.6

These probes were first subjected to S1 digestion individually to test for enzymatic activity. The background oxidation of IC-098 was studied in S1 buffer, and unfortunately there was an additional peak centred around -100 mV. Upon treatment with S1 nuclease, the impurity peak did not change oxidation potential or increase in current, while the peak centred

around 180 mV shifted roughly 80 mV and increased the current generated, indicating successful digestion had occurred despite contamination with impurities. Due to the Cp* derivative's issues with purification, this was not entirely surprising. Additionally, the mass return of these probes was significantly lower than other substituted oligonucleotides, further indicating probe instability.

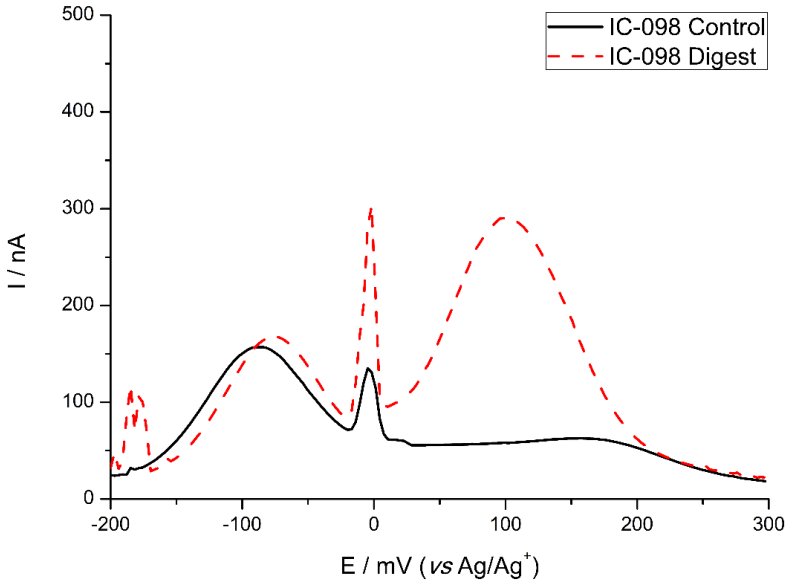


Figure 4.16 IC-098 probe S1 digestion. Probe (3 μM), pH 4.5 NaOAc buffer. Negative control or S1 (10 U). 37 °C, 30 minutes in tube, Autolab detection.

The *N. gonorrhoeae* probe was also subjected to the S1 digestion protocol, displaying full digestion in 30 minutes (Figure 4.17). Despite similar background currents, both the NG and IC probes were shown to produce significantly reduced signal upon digestion compared to the CT derivative (Figure 4.18).

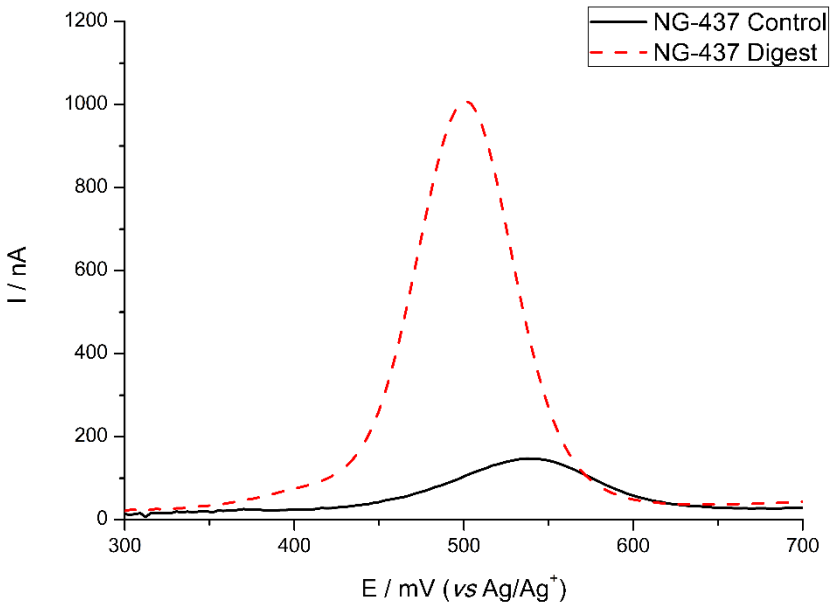


Figure 4.17 NG-437 probe S1 digestion. Probe (3 µM), pH 4.5 NaOAc buffer. Negative control or S1 (10 U). 37 °C, 30 minutes in tube, Autolab detection.

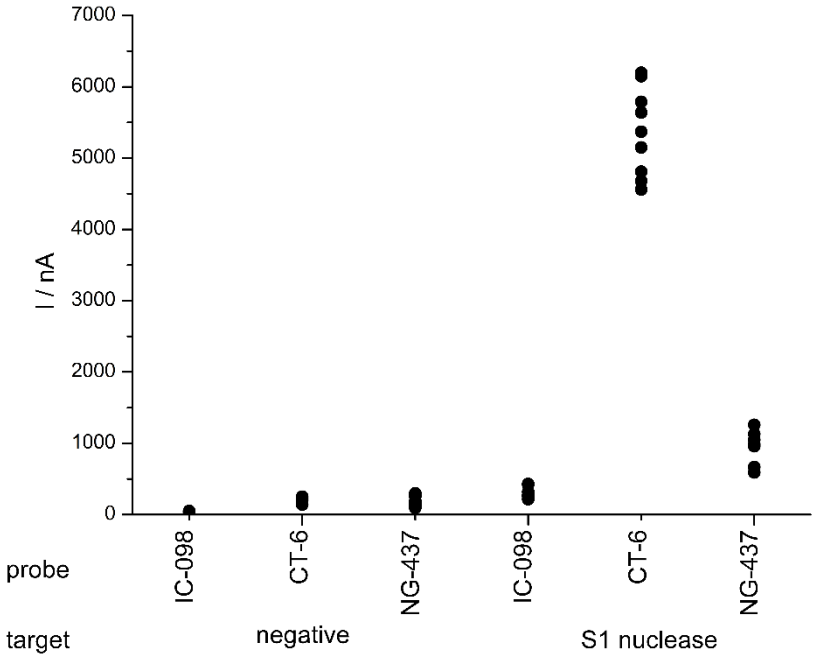


Figure 4.18 IC/CT/NG probes, S1 individual digestions. Probe (3 µM), pH 4.5 NaOAc buffer. Negative control or S1 (10 U). 37 °C, 30 minutes in tube, Autolab detection.

The probes were then detected in the same solution to test for multiplex capability. The probes were able to fully multiplex together, clearly showing the three different labels, despite the impurity from the IC-098 probe (Figure 4.19).

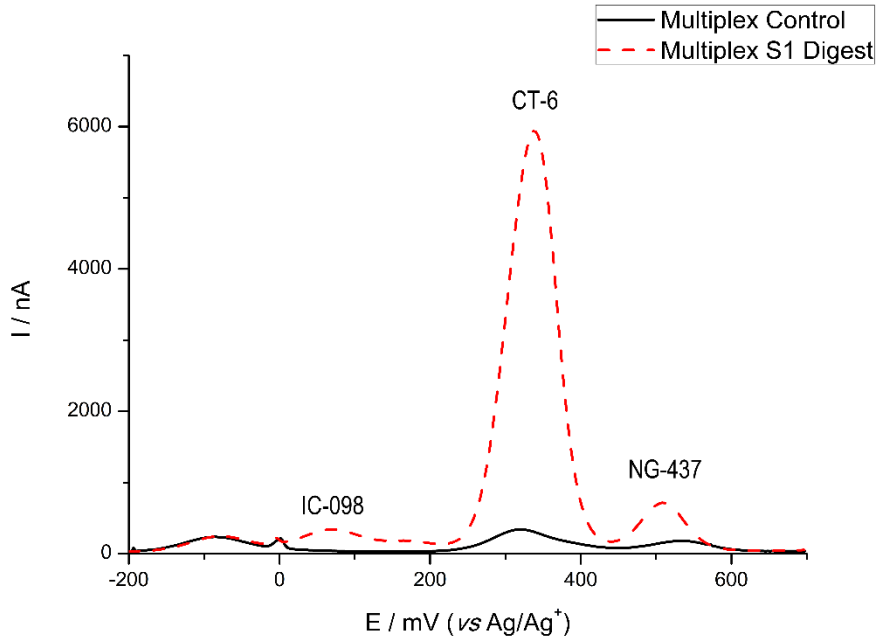


Figure 4.19 IC/CT/NG probes combined S1 digestion. IC-098 (3 µM) CT-6 (3 µM), NG-437 (3 µM), pH 4.5 NaOAc buffer. Negative control or S1 (10 U). 37 °C, 30 minutes in tube, Autolab detection.

From visual inspection, it is difficult to immediately see the IC and NG probes, however upon peak search and analysis through individual value plot (Figure 4.20), both a shift in oxidation potential and an increase in signal are observed for each probe. Crucially, combining all three probes into a single mixture did not inhibit the digestion of any of the probes, with all three able to undergo digestion with S1 nuclease. The signal of each probe is relatively unaffected by combining all three probes into a single solution, with only a slight reduction in current generated for IC-098 upon multiplex detection.

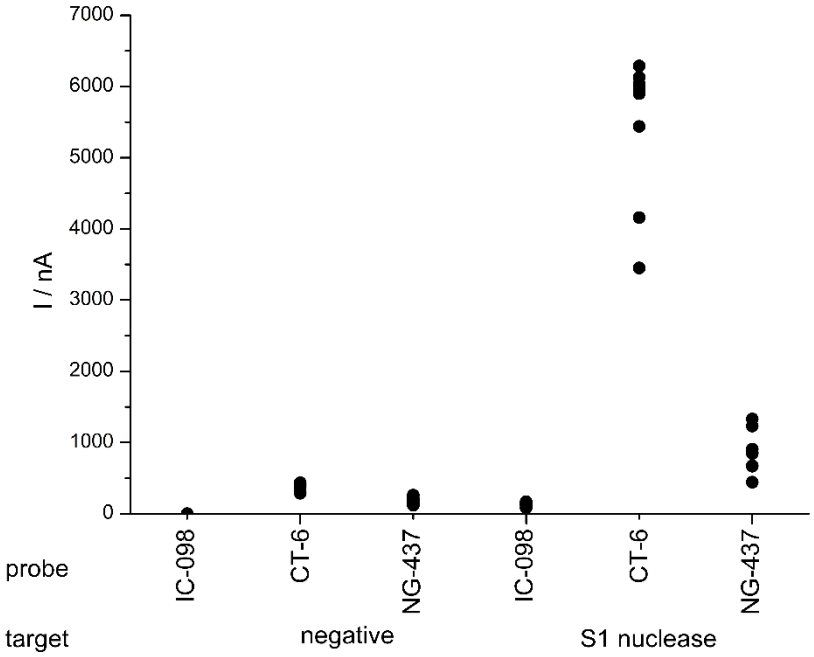


Figure 4.20 IC/CT/NG probes combined S1 digestion. IC-098 (3 µM), CT-6 (3 µM), NG-437 (3 µM), pH 4.5 NaOAc buffer. Negative control or S1 (10 U). 37 °C, 30 minutes in tube, Autolab detection.

The probes were then assessed for suitability in a double-stranded assay. The probes were first digested with amplified DNA in separate assays to individually assess their reactivity towards T7 exonuclease.

Unfortunately, the IC-098 probe did not digest showing no change in either E_{ox} or current after treatment with T7 and IC amplicons. This could be due to the proximity of the label to the start of the sequence, however the IC-1 probe was able to undergo digestion, which has the same labelling pattern, differing in the ferrocene label used. Alternatively, the size of the di-098 label could interfere with the hybridisation or enzymatic activity of T7.

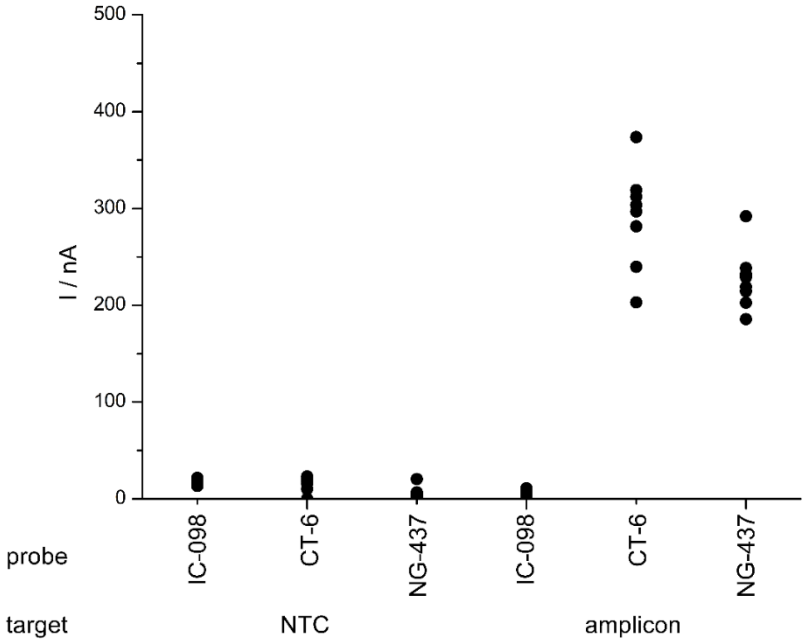


Figure 4.21 IC/CT/NG probes, double stranded individual digestions. Detection mix: Probe (6 μM), detection buffer, T7 exo (16 U). DNA mix: NTC PCR mix (20 μL), or IC PCR (10 pg) mix (20 μL), or CT PCR (100 IFUs), or NG PCR (100 nmoles). 37 °C, 3 minutes in tube, Autolab detection.

Due to the inactivity of IC-098 in target assays, it was eliminated from the multiplex assay. To this extent, a simultaneous CT/NG PCR was conducted and detected using a CT/NG combined detection mix. Pleasingly, this showed both probes underwent digestion in the target assay, with minimal detectable backgrounds. The peaks can clearly be distinguished from NTCs (Figure 4.22). Analysis of the results clearly shows both probes have been successfully digested by T7 exonuclease in the presence of amplified DNA, displaying both a shift in E_{ox} and an increase in current.

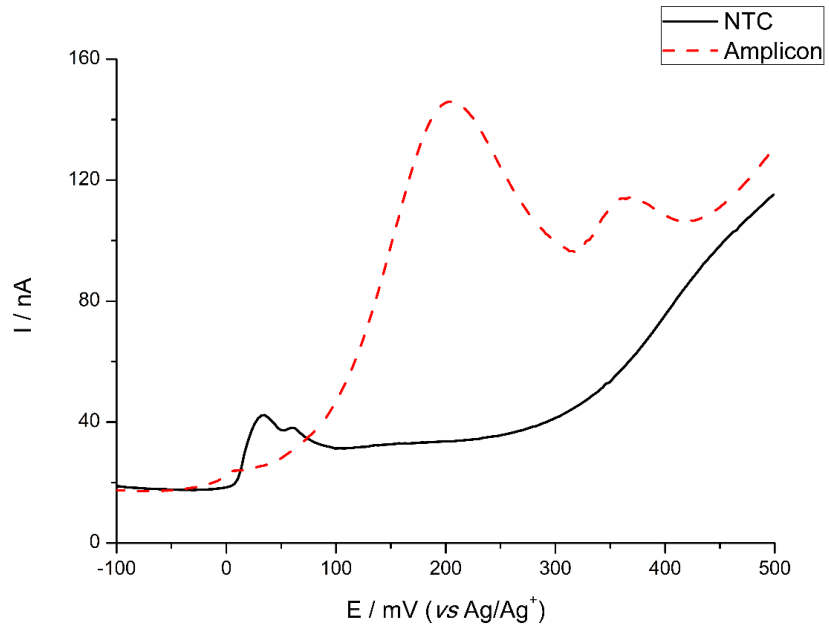


Figure 4.22 Voltammogram of CT/NG multiplex digestion. Detection mix: CT-6 (6 µM), NG-437 (6 µM), detection buffer, T7 exo (16 U). DNA mix: NTC PCR mix (20 µL), or combined CT (100 IFUs)/NG (100 nmoles) PCR. 37 °C, 3 minutes in tube, Autolab detection.

Despite their similar performance in individual assays, the NG probe resulted in slightly reduced signal compared to the CT probe (Figure 4.23). The CT and NG probes both showed large increase in signal compared to the NTCs.

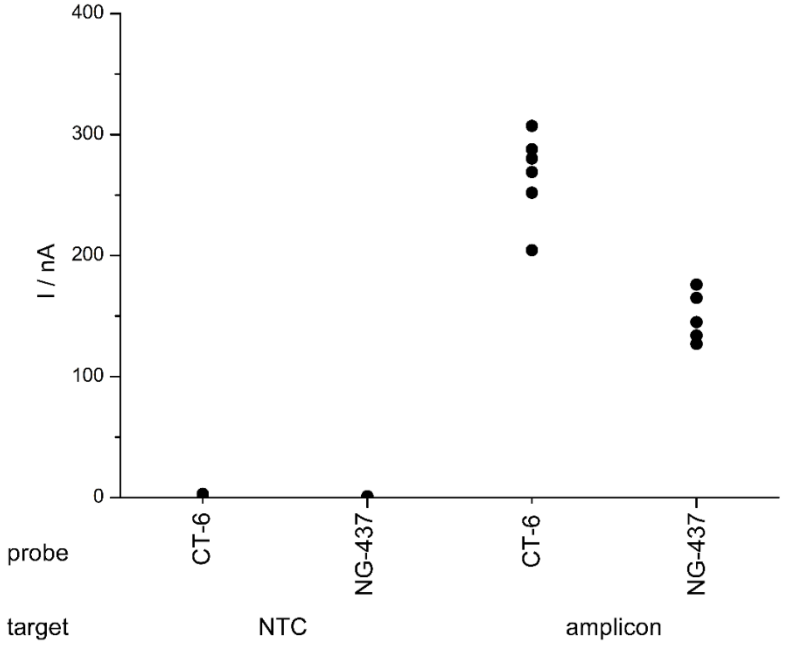


Figure 4.23 CT/NG multiplex digestion individual value plot. Detection mix: CT-6 (6 µM), NG-437 (6 µM), detection buffer, T7 exo (16 U). DNA mix: NTC PCR mix (20 µL), or combined CT (100 IFUs)/NG (100 nmoles) PCR. 37 °C, 3 minutes in tube, Autolab detection.

These results show the internally labelled probes are suitable for use in a duplex detection assay. Disappointingly, one of the labels chosen did not digest under the conditions studied, however combined with other results this can be attributed to the properties of the ferrocene label itself being incompatible, rather than due to the internal labelling strategy. Further studies of ferrocene derivatives incorporating electron-donating substituents may allow the identification of multiplex labels.

4.4 Lambda-exonuclease digestions

The digestion of dsDNA with T7 exonuclease is very quick, being complete in only 3 minutes. It is believed its mode of action involves "jumps" on and off the duplex, digesting small di-, tri- and tetranucleosides at a time. Once T7 has digested the target strand to a certain degree, the duplex will no longer be stable due to its reduced T_m, and will denature at the detection temperature of 37 °C. This hypothesis is supported with the results shown by CT-13, which has a labelled uridine residue at the 3'-end, on the 23rd base pair. The mechanism of lambda exonuclease (λ or λ-exo) is a threading action from 5' to 3', whereby the enzyme threads DNA through the active site, releasing mononucleotides and ssDNA.[188] It is 350 times more active on dsDNA than ssDNA and the substrate requires 5'-phosphorylated DNA[189]

The traditional 5'-labelled probes are unsuitable substrates for λ-exo due to the bulky 5'-ferrocene label and lack of 5'-phosphate. After design of the new internal labelling, it was thought the new probes may be suitable substrates for λ-exo.

Probe	Sequence
CT-13	5'-[p]-CTG TCC GCT GGT TCT TCC TTA C<u>U^{d275}</u>-[p]-3'
CT-6	5'-[p]-C<u>U^{d275}</u>G TCC GCT GGT TCT TCC TTA CT-[p]-3'
CT-dC	5'-**Z**- CTG TCC GCT GGT TCT TCC TTA CT-[p]-3'

Table 4.7 Lambda exonuclease probes

Initial studies conducted by AG on CT-6 and CT-dC indicated that while λ-exo successfully digested PCR amplicons, as demonstrated by gel electrophoresis, neither the 5'- or internally labelled probes digested to a significant level in 60 minutes.[190] As both of these probes were shown to digest after treatment with cDNA and T7 exonuclease, it was believed the mid-strand ferrocene in CT-6 was blocking the action of λ-exo. This was later discovered, however, to be due to the length of the 3'-overhang generated in PCR, as overhangs greater than 100 base pairs in length inhibit the digestion reaction 30-fold.[191]

After synthesis of probe CT-13, the experiments were repeated with synthetic target rather than amplicon, showing that previous results indicating CT-6 was inactive were masking the true reactivity of the probe. The runs were repeated, showing that CT-6 and ST cDNA do indeed digest in the presence of λ-exo, while probes CT-dC and CT-13 are completely inactive (Figure 4.24).

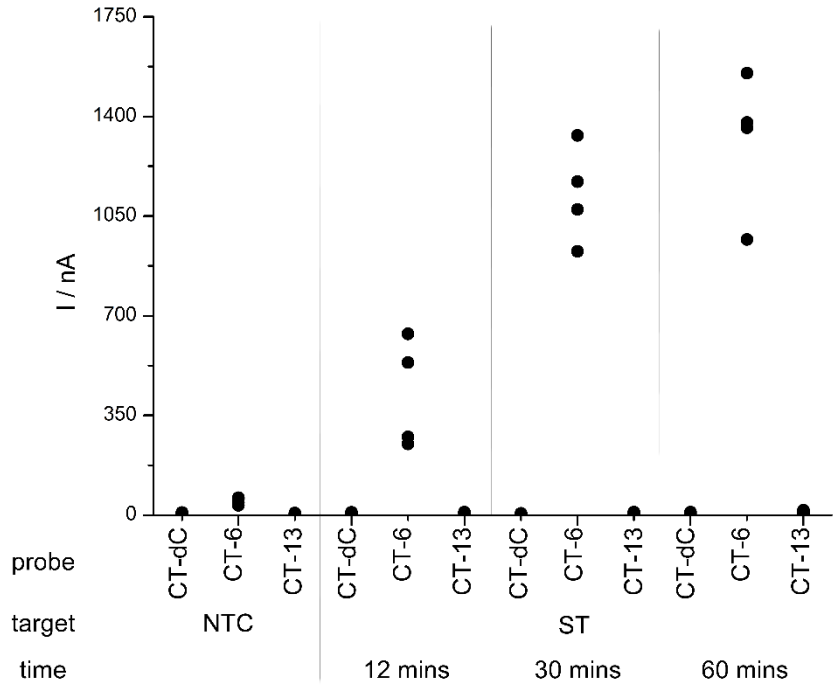


Figure 4.24 CT probes, λ-exonuclease susceptibility testing. Probe (6 μM), lambda buffer, λ-exonuclease (5 U), synthetic target mix (12 μM),

λ-exo has slower reactivity than T7 exo, which is seen in the growth of signal over a 60-minute period. The signal generated, however, is far superior to that of the T7 assay, with 12-minute digestion with λ-exo offering similar signal to the 3-minute digestion of T7 exo. The enzyme did, however, generate higher signals, which can be attributed to the known release of mono-nucleotides and an intact single-stranded DNA. The mono-nucleotides released are likely to have superior signal compared to the dimers and trimers that are thought to be released with T7 exonuclease.

The extension of the redox active nucleotides into probes which are selective for particular enzymes could allow for improved diagnostic assays, or use of labels which were previously unsuitable with T7 exonuclease. A two-enzyme system could allow for logic gate style functionality if suitable buffer systems can be created.

4.5 Conclusion

This chapter has discussed the application of internally labelled oligonucleotides in double-stranded DNA assays. The internally labelled probes were shown to undergo digestion in the same fashion to 5'-labelled control probes without any increase in signal, as was observed for the S1 digestions. Integration of the probes onto the diagnostic platform proved successful, however no significant performance increase compared to the 5'-labelled controls was observed.

The internal multiplex labels were incorporated into suitable probe sequences and used to detect a CT/NG combined PCR in a single assay.

Along with the synthesis of a number of sequence permutations, the use of multiplex system offered some insight into the necessary ordering of labelled oligonucleotides to allow for digestion. These results indicate labels must be at least three base pairs away from each other to allow for optimal digestion. It is here that the mass spectrometry labels would be particularly useful, to identify intact fragments to determine how much of the oligonucleotide is being successfully digested. However, studies conducted with oligonucleotide digests did not offer information as due to high dilution and high salt concentration. Further development of these labels, in combination with de-salting columns to improve resolution, could provide valuable information on sequence selection.

Due to the additional synthesis required for the internally labelled probes, studies conducted in this thesis suggest that they are not worth pursuing for use with T7 to offer an improved digestion protocol.

The new internally labelled probes allowed for use of λ-exonuclease, which has been previously unsuitable for use with the 5'-labelled probes. The digested probe gave higher signal than its T7 equivalent assay in a similar timeframe in an un-optimised system. Further development of this assay could allow for the incorporation of labels that have been unsuitable (*e.g.* IC-098 if the impurity issue can be overcome) and improve the multiplex properties of the system. Additionally, λ-exonuclease is a cheaper than T7-exonuclease, which upon scale up and commercialisation of the assay could save AG on costs.

4.6 Summary

The probes described in this thesis were subjected to digestion with various nucleases to assess any potential improvements in electrochemical current that could be achieved. Initial results of ssDNA digestions with S1 nuclease indicated that improved digestion could be achieved through additional ferrocene substitution. From the data collected, it has been shown that there is no real benefit to introducing more than three substituted sites due to the increased material required for oligonucleotide synthesis, as these did not display improved digestion, which would not outweigh the additional cost of custom oligonucleotide.

The incorporation of di-ferrocenes synthesised *via* reductive amination offered dramatically improved sensitivity over mono-ferrocene probes, however concurrent with other S1 nuclease studies, there was no benefit to including more than three modified nucleotides in the sequence.

The move to dsDNA digestions showed promising results with, showing that some increase in electrochemical signal was possible. The probes were shown to undergo digestion in the presence of T7 exonuclease and complementary DNA which had been synthesised by both solid phase and enzymatic methods. The use of synthetic showed promise, indicating increased current was indeed possible, however the move to cDNA generated *via* PCR showed no difference between any of the probes synthesised. Given the additional chemical synthesis required for the modified oligonucleotides, it would not be a commercially viable route unless significant improvements could be achieved.

The probes were subjected to digestion with lambda exonuclease, which showed promising results, showing that 5'-labelled probes were not suitable for use with this nuclease, however the introduction of internal labels allowed a vastly improved digestion. The system has not yet been optimised, and as such further work would focus on finding the optimal buffer solution and additives, as well as aiming to reduce the digestion times required.

Further work arising from this thesis would include variation on labelling position of pyrimidine. Current 5-position labelling is located in the major groove section, however to compare any stabilising or destabilising effects of substitution, synthesis of the equivalent 6-substituted nucleoside would allow for analysis of the effects of the chemical modification. UV-melt analysis would allow for determination of T_m, indicative of duplex stability, which would further inform optimal labelling position on both the nucleobase itself, and the labels' position within the oligonucleotide.

5. **Experimental**

5.1. General considerations

5.1.1. Equipment

^{1}H, ^{13}C, and ^{31}P nuclear magnetic resonance (NMR) spectra were recorded on a Bruker Avance 300 MHz spectrometer or an Agilent Technologies 500 MHz spectrometer at 298 K unless otherwise stated. Chemical shifts are reported in parts *per* million downfield from $Si(CH_3)_4$ and are referenced to residual solvent peaks,[192] and the multiplicity (s = singlet, d = doublet, dd = doublet of doublets, dt = doublet of triplets, t = triplet, td = triplet of doublets, tt= triplet of triplets, q = quartet, p = pentet, h = hextet, m = multiplet, app. = apparent), coupling constant(s) (in hertz), and integration are reported. Electrospray ionisation high resolution time-of-flight mass spectrometry (HRMS) were recorded on a Bruker micrOTOF spectrometer. Infrared (IR) spectra were recorded on a PerkinElmer 1600 Fourier transform IR spectrometer with selected absorbencies quoted as wavenumber (ν (cm^{-1})). PCR tubes were heated using an MJ Research PTC-200 Thermal Cycler. Electrochemical analysis was performed on a Metrohm Autolab PGSTAT30 potentiostat controlled by a personal computer running General Purpose Electrochemical System (GPES) software. Samples of 24 μL were applied to electrodes and were analysed using differential pulse voltammetry (modulation = 0.04 s, interval = 0.1 s, step potential = 3mV, modulation amplitude = 50 mV), the voltammograms were baseline corrected using the software's "baseline correction" function, and peak current was measured through the software's "peak search" function. io® electrochemical analysis was performed on an Atlas Genetics io with external Peltier-potentiostat using differential pulse voltammetry (modulation = 0.04 s, interval = 0.06 s, step potential = 3 mV, modulation amplitude = 50 mV), and the voltammograms were baseline corrected and peak current was calculated in the corresponding range using AG's peak search algorithm.

5.1.2. Reagents

Silica gel chromatography was performed using 60 Å, 200-400 mesh silica gel purchased from Sigma-Aldrich. Analytical thin-layer chromatography (TLC) was performed using aluminium-backed plates coated with Alugram® SIL G/UV_{254} purchased from Fischer and visualised by UV light (254 nm), and/or $KMnO_4$ staining. Screen printed electrodes (carbon working and counter, Ag/Ag^+ reference) were donated by Atlas Genetics (Trowbridge, UK).

Reactions were performed using oven-dried glassware. All reactions used solvents and reagents as obtained from Sigma Aldrich without further purification, unless otherwise stated. Petrol refers to petroleum ether (40-60 °C). All temperatures quoted are external. 5-iodo-2'-deoxyuridine was purchased from Carbosynth (Cambridge, UK). Di-*tert*-butyl

dicarbonate was purchased from Fluorochem (Hadfield, UK). 6-(bis(ferrocenylmethyl)amino)hexan-1-ol, 6-(bis((chloroferrocenyl)methyl)amino) hexan-1-ol, and 6-(bis(ruthenocenenylmethyl)amino)hexan-1-ol were donated by Atlas Genetics (Trowbridge, UK). Oligonucleotides were synthesised by ATDbio (Southampton, UK) and used as supplied 100 μM aqueous solutions. S1 nuclease (100 U/μL) was purchased from ThermoFisher Scientific and diluted prior to use (1 U/μL in molecular biology grade H_2O). S1 5X reaction buffer was used as supplied by ThermoFisher Scientific. dUTPs were purchased from Bioline (London, UK). T7 exonuclease was purchased from ThermoFisher. *Taq*-B and UNG were purchased from Enzymatics (MA, USA). Lambda exonuclease emzyme and buffer were purchased from New England BioLabs (MA, USA).

5.2. Biological Procedures

5.2.1. Primer sequences

Oligonucleotide	Sequence
IC Forward	5' – TCG CTG TCG GGA AGT TTG GTT GAA – 3'
IC Reverse	5' – AGG CCT GAA CTG GGA ATC CTT TG – 3'
CT Forward	5' – GTT TGG ACA CTA GTC AGC ATC AAG CTA GG – 3'
CT Reverse Mod	5' – A*G*A* T*TC CAG AGG CAA TGC CAA AGA AA – 3'
NG2 Forward	5' – ACG CAA ACG GAG GTC TTA CGG ATT TAG – 3'
NG2 Reverse	5' – CGT TGG CGC AAT TTC CAT ATA GTC CTG – 3'
IC Synthetic Target	5' – GTG AAA ATC GGT AAC GTC TTT AGG AAA GGG ATC GTG CTC TCC GGT TCA ACC AAA C – 3'
CT Synthetic Target	5' – GCC AAA GAA AAA AGT AAG GAA GAA CCA GCG GAC AGG ACG GCG ATG CTT CCT TTT A – 3'

Table 5.1 Primer and target sequences. * indicates phosphorthioate ($P(S)O_3$) linkages

5.2.2. Buffer solutions

Elution Buffer

EDTA (500 mM) in pH 8.9 Trizma HCl (496 mM) buffer.

50x PCR Buffer

KCl 2.47 M, Trehalose (1.01 M) in pH 9.0 Trizma HCl (496mM) buffer.

Detection surfactant solution

Amino-butyric-acid (356 mM), Trehalose (0.971 M) and Triton X-305 (0.25% v/v) in water.

T7 Diluent

Dithiothreitol (3.48 mM), ethylenediaminetetraacetic acid (10mM) and BSA (2% w/v) in pH 6.4 K_3PO_4 buffer (216 mM).

S1 buffer used as supplied from ThermoFisher.

Lambda buffer used as supplied from New England BioLabs.

5.2.3. Enzymatic assay protocols & mixtures

5.2.3.1. S1 Digestions

The appropriate probe was chosen and mixed according to Tables 5.2–5.5. A master solution was made, with the addition of S1 nuclease (if applicable) the final addition. The solutions were shaken by hand, and aliquoted in 30 µL portions into PCR tubes. The tubes were heated to 37 °C for 30 minutes and stored at 4 °C prior to detection for a maximum of 15 minutes.

Component	Stock Conc.	Final Conc.	Required Volume
Buffer	-	-	6
Probe	100 µM	6 µM	0.9
Water	-	-	13.1
S1 Nuclease	1 U/µL	10 U/rxn	10

Table 5.2 S1 Digestions (3 µM probe)

Component	Stock Conc.	Final Conc.	Required Volume
Buffer	-	-	6
Probe	100 µM	6 µM	0.9
Water	-	-	23.1

Table 5.3 S1 Nuclease negative controls (3 µM probe)

Component	Stock Conc.	Final Conc.	Required Volume
Buffer	-	-	6
Probe	100 µM	6 µM	1.8
Water	-	-	12.2
S1 Nuclease	1 U/µL	10 U/rxn	10

Table 5.4 S1 Digestions (6 µM probe)

Component	Stock Conc.	Final Conc.	Required Volume
Buffer	-	-	6
Probe	100 µM	6 µM	1.8
Water	-	-	22.2

Table 5.5 S1 Nuclease negative controls (6 µM probe)

5.2.3.2. PCRs and synthetic target mixes

The appropriate PCR mixture was made into a master mix according to Tables 5.6–5.9. For no template controls, target DNA was replaced with MBG water.

The solutions were vortexed for 10 s, aliquoted into PCR tubes (64 µL), and thermally cycled. After thermal cycling, each tube was combined into a bijou tube and vortexed for 10 s. The amplicons were stored at –20 °C for up to 3 months prior to detection.

Protocol: 37 °C for 120 s, 94 °C for 180 s, then cycle x40 (94 °C (2 °C/s) for 5 s, 65 °C (2 °C/s) for 10 s) then store at 8 °C.

Component	Stock conc.	Final conc.	Required volume (µL)
50x PCR buffer	-	-	1.28
dUTP mix	50 mM	1 mM	1.28
UNG	50 U/µL	0.05 UµL	0.06
*Taq*B	25 U/µL	0.10 U/µL	0.26
$MgCl_2$	1.0 M	5 mM	0.32
IC Forward	10 µM	0.04 µM	0.26
IC Reverse	100 µM	0.3 µM	0.19
Brij 58	5% w/w	0.02%	0.26
MBG Water	-	-	1.18
IC DNA	100 pg/µL	10 pg/rxn	1
Elution buffer	-	-	53.91

Table 5.6 Internal Control PCR Primer mixes

Component	Stock conc.	Final conc.	Required volume (µL)
50x PCR buffer	-	-	1.28
dUTP mix	50 mM	1 mM	1.28
UNG	50 U/µL	0.05 UµL	0.06
*Taq*B	25 U/µL	0.10 U/µL	0.26
$MgCl_2$	1.0 M	5 mM	0.32
CT Forward	100 µM	0.40 µM	0.26
CT Reverse Mod	100 µM	0.60 µM	0.38
Brij 58	5% w/w (H_2O)	0.02%	0.26
MBG Water	-	-	1.18
CT DNA	1000 IFU/µL	100 IFU/rxn	1
Elution buffer	-	-	53.72

Table 5.7 *Chlamydia trachomatis* PCR Primer mixes

Component	Stock conc.	Final conc.	Required volume (µL)
50x PCR buffer	-	-	1.28
dUTP mix	50 mM	1 mM	1.28
UNG	50 U/µL	0.05 UµL	0.06
*Taq*B	25 U/µL	0.10 U/µL	0.26
$MgCl_2$	1.0 M	3.5 mM	0.22
NG2 Forward	100 µM	0.45 µM	5.76
NG2 Reverse	100 µM	0.45 µM	5.76
Brij 58	5% w/w (H_2O)	0.02%	0.26
MBG Water	-	-	0.27
NG DNA	100 nM	1.56 nM	1
Elution Buffer	-	-	57.72

Table 5.8 *Nerissa gonorrhoeae* PCR Primer mixes

Component	Stock conc.	Final conc.	Required volume (µL)
50x PCR buffer	-	-	1.28
dUTP mix	50 mM	1 mM	1.28
UNG	50 U/µL	0.05 UµL	0.06
*Taq*B	25 U/µL	0.10 U/µL	0.26
$MgCl_2$	1.0 M	5 mM	0.32
CT Forward	100 µM	0.40 µM	0.25
CT Reverse Mod	100 µM	0.60 µM	0.38
NG2 Forward	100 µM	0.45 µM	0.29
NG2 Reverse	100 µM	0.45 µM	0.29
Brij 58	5% w/w	0.02%	0.26
MBG Water	-	-	0.27
CT DNA	100 IFU/µL	100 IFU/rxn	1
NG DNA	100 nM	1.56 nM	1
Elution Buffer	-	-	53.06

Table 5.9 NG/CT multiplex PCR primer mix

Component	Stock Conc.	Final Conc.	Required volume (µL)
Synthetic Target DNA	100 µM	12 µM	2.76
MgCl	1.0 M	5 mM	0.1
50x PCR buffer	-	-	0.4
Elution buffer	-	-	16.74

Table 5.10 Synthetic target mix

5.2.3.3. Double-stranded digestions

The appropriate detection mix was made into a master mix and the appropriate procedure was followed.

Digestion in tube

The detection mix was aliquoted into PCR tubes and the appropriate DNA mix (amplicon or synthetic target) (20 µL) was added. The tubes were vortexed for 5 s, centrifuged for 5 s, and heated to 37 °C for 3 minutes in a thermocycler. The mixtures were stored at 4 °C for up to 15 minutes prior to detection.

Digestion "on sub-circuits"

The detection mix was aliquoted into PCR tubes and the appropriate DNA amplicon (20 µL) was added. The tubes were vortexed for 5 s, centrifuged for 5 s, and then applied to the electrode sub-circuit, inserted into the clamping rig, then heated to 37 °C for 3 minutes, followed by immediate detection at 37 °C.

Digestion "dried-down"

The detection mix was applied to bare sub-circuits and heated to 45 °C at 10% relative humidity for 10 minutes. Adhesive and electrodes were used to seal the sub-circuit, and were used the same day. To the sub-circuit was applied amplicon (24 µL), mixing by adding to the first stop and rest position three times. The sub-circuit was inserted into the clamping rig, heated to 37 °C for 3 minutes, followed by immediate detection at 37 °C.

Component	Stock Conc.	Final Conc.	Required volume (µL)
Detection surfactant solution	-	-	0.29
T7 diluent	-	-	0.45
T7 exonuclease	50 U/µL	0.825 U/µL	0.33
Probe	100 µM	6 µM	1.45
Water	-	-	1.45

Table 5.11 Single probe (6 µM) detection mix

Component	Stock Conc.	Final Conc.	Required volume (µL)
Detection surfactant solution	-	-	0.29
T7 diluent	-	-	0.45
T7 exonuclease	50 U/µL	0.825 U/µL	0.33
Probe	100 µM	12 µM	2.89

Table 5.12 Single probe (12 µM) detection mix

Component	Stock Conc.	Final Conc.	Required volume (µL)
Detection surfactant solution	-	-	0.29
T7 diluent	-	-	0.45
T7 exonuclease	50 U/µL	0.825 U/µL	0.33
CT Probe	100 µM	6 µM	1.45
NG2 Probe	100 µM	6 µM	1.45

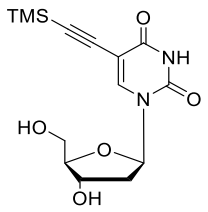
Table 5.13 CT/NG Multiplex detection mix

Component	Stock Conc.	Final Conc.	Required volume (µL)
1,4-dithiothreitol	34.8 mM	0.67 µM	0.513
Bovine serum albumin	0.925%	0.02%	0.513
Detection surfactant solution	-	-	0.28
Lambda buffer	-	-	1.9
Lambda exonuclease	5 U/µL	5 U/rxn	1
Probe	100 µM	6 µM	1.78

Table 5.14 Lambda exonuclease digestion mix

5.3. Chemical procedures and characterisation

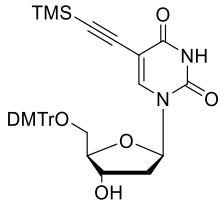
5-(trimethylsilyl)ethynyl-2'-deoxyuridine 25[141]



To an oven-dried round bottom flask was added 5-iodo-2'-deoxyuridine (2.48 g, 7 mmol) and dissolved in anhydrous MeCN (87 mL) and triethylamine (87 mL). Argon was bubbled through the solution for 10 minutes with stirring. To the colourless solution was added bis(triphenylphosphine)palladium(II) chloride (98 mg, 0.14 mmol) and copper (I) iodide (67 mg, 0.35 mmol) and stirred for 10 minutes under argon. The yellow solution was treated with trimethylsilyl acetylene (3.95 mL, 28 mmol) and heated to 50 °C for 4 hours. The reaction was cooled to rt and solvent was removed to give a thick gum, which was purified *via* flash chromatography, eluting with $CHCl_3$/MeOH (19:1) to yield the title compound as an off-white foam (1.98 g, 86%).

MP 160–162 °C (lit. 162–165 °C).[142] **^{1}H NMR** (500 MHz, CD_3OD) δ 8.29 (s, 1H, Ar*H*), 6.19 (t, *J* = 6.5 Hz, 1H, 1'-*H*), 4.36 (dt, *J* = 6.5, 3.6 Hz, 1H, 3'-*H*), 3.89 (q, *J* = 3.3 Hz, 1H, 4'-*H*), 3.78 (dd, *J* = 12.0, 3.3 Hz, 1H, 5'-*H*), 3.70 (dd, *J* = 12.0, 3.3 Hz, 1H, 5'-*H*), 2.27 (ddd, *J* = 13.6, 6.5, 3.6 Hz, 1H, 2'-*H*), 2.23–2.14 (m, 1H, 2'-*H*), 0.16 (s, 9H, 3 x C*H*$_3$). **^{13}C NMR** (126 MHz, CD_3OD) δ 163.1 (*C*O), 150.1 (Ar*C*), 144.5 (Ar*C*), 99.1 (Ar*C*), 97.6 (*C*), 95.8 (*C*), 87.7 (4'-*C*). 85.6 (1'-*C*), 70.5 (3'-*C*), 61.1 (5'-*C*), 40.3 (2'-*C*), -1.5 (3 x CH_3).

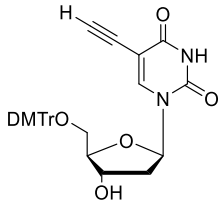
5-(trimethylsilyl)ethynyl-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridine 26[143]



To a stirred solution of **25** (670 mg, 2.0 mmol) in freshly distilled pyridine (8 mL) was added DMTrCl (746 mg, 2.2 mmol). The orange solution was allowed to stir at rt for 2 hrs. The reaction was cooled to 0 °C and quenched with MeOH (1 mL). The solution was concentrated *in vacuo* to an oil and dissolved in CH_2Cl_2 (20 mL) and washed with $NaHCO_{3(sat.)(aq.)}$ (3 x 20 mL) and $brine_{(sat.)}$ (20 mL) and the organics were dried over $MgSO_4$, and concentrated *in vacuo.* The crude solid was purified *via* silica gel chromatography (CH_2Cl_2/MeOH/TEA 99:1:0.5) to yield a colourless foam (755 mg, 60%).

MP 109–110 °C. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.98 (s, 1H, Ar*H*), 7.46–7.40 (m, 2H, Ar*H*), 7.35–7.31 (m, 4H, Ar*H*), 7.28 (t, *J* = 7.8 Hz, 2H, Ar*H*), 7.22–7.16 (m, 1H, Ar*H*), 6.89–6.68 (m, 4H, Ar*H*), 6.29 (dd, 8.0, 5.8 Hz, 1H, 1'-*H*), 4.41–4.43 (m, 1H, 3'-*H*), 4.07 (q, *J* = 3.4 Hz, 1H, 4'-*H*), 3.77 (s, 3H, OC*H*$_3$), 3.76 (s, 3H, OC*H*$_3$), 3.38 (dd, *J* = 10.6, 3.4 Hz, 1H, 5'-*H*), 3.28 (dd, *J* = 10.6, 3.4 Hz, 1H, 5'-*H*), 2.46 (ddd, *J* = 13.5, 5.8, 2.6 Hz, 1H, 2'-*H*), 2.17 (ddd, *J* = 13.5, 8.0, 5.8 Hz, 1H, 2'-*H*), -0.01 (s, 9H, 3 x C*H*$_3$). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 162.6 (*C*O), 158.6 (Ar*C*), 150.1 (CO), 144.5 (Ar*C*), 142.7 (Ar*C*), 135.7 (Ar*C*), 130.0 (Ar*C*), 128.1 (Ar*C*), 128.0 (Ar*C*), 127.0 (Ar*C*), 113.4 (Ar*C*), 100.6 (Ar*C*), 99.2 (Ar*C*), 95.5 (*C*), 86.9 (4'-*C*), 86.5 (O*C*Ar$_3$), 85.7 (1'-*C*), 72.2 (3'-*C*), 63.6 (5'-*C*), 55.3 (2 x O*C*H$_3$), 41.5 (2'-*C*), -0.2 (*C*H$_3$).

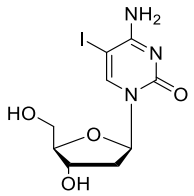
5-ethynyl-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridine 27[193]



To a stirred slurry of **26** (5.18 g, 8.3 mmol) in THF (9 mL) was added TBAF (1.0 M in THF) (12.4 mL, 12.4 mmol) in one portion and allowed to stir at rt overnight. The solution was concentrated and purified *via* silica gel chromatography (EtOAc) to yield the title compound as a colourless foam (3.14 g, 58%).

MP 123–125 °C. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.08 (s, 1H, Ar*H*), 7.44–7.38 (m, 2H, Ar*H*), 7.32 (dd, *J* = 8.9, 1.8 Hz, 4H, Ar*H*), 7.29 (t, *J* = 7.7 Hz, 2H, Ar*H*), 7.24–7.17 (m, 1H, Ar*H*), 6.84 (dd, *J* = 8.9, 1.8 Hz, 4H, Ar*H*), 6.28 (dd, *J* = 7.4, 5.8 Hz, 1H, 1'-*H*), 4.58–4.51 (m, 1H, 3'-*H*), 4.12–4.07 (m, 1H, 4'-*H*), 3.78 (s, 6H, 2 x C*H*$_3$), 3.41–3.36 (m, 2H, 5'-*H*), 2.89 (s, 1H, C*H*), 2.52 (ddd, *J* = 13.7, 5.8, 2.5 Hz, 2H, 2'-*H*), 2.32–2.22 (m, 1H, 2'-*H*). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 161.3 (CO), 158.6 (Ar*C*), 149.0 (CO), 144.4 (Ar*C*), 143.6 (Ar*C*), 135.4 (Ar*C*), 130.0 (Ar*C*), 128.0 (Ar*C*), 127.9 (Ar*C*), 127.0 (Ar*C*), 113.3 (Ar*C*), 99.2 (Ar*C*), 87.1 (4'-*C*), 86.4 (O*C*Ar$_3$), 85.7 (1'-*C*), 82.0 (*C*), 74.0 (*C*), 72.2 (3'-*C*), 63.3 (5'-*C*), 55.2 (2 x O*C*H$_3$), 41.4 (2'-*H*).

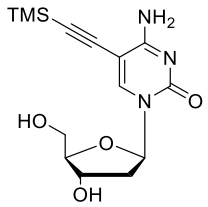
5-iodo-2'-deoxycytidine 29[146]



To a round-bottom flask fitted with magnetic stirrer bar was added 2'-deoxycytidine (6.80 g, 30 mmol), iodic acid (2.70 g, 15.4 mmol), and iodine (4.50 g, 17.7 mmol), and the solids were dissolved in acetic acid (24 mL) CCl_4 (6 mL) and H_2O (9 mL) and the solution was heated to 40 °C for 90 minutes. The reaction was cooled to rt, diluted with H_2O (50 mL) and filtered. The filtrate was concentrated *in vacuo* and co-evaporated with MeOH (3 x 20 mL). The solid was dissolved in water, adjusted to pH 10 and concentrated *in vacuo*. The solid was dissolved in the minimum amount of hot water and allowed to recrystallise at 4 °C overnight to yield the title compound as a white powder (3.25 g, 31%).

MP 137–139 °C decomp. (lit. 133–135 (decomp.)).[146] **^{1}H NMR** (500 MHz, $(CD_3)_2SO$) δ 8.28 (s, 1H, Ar*H*), 6.07 (t, *J* = 6.4 Hz, 1H, 1'-*H*), 4.21 (dt, *J* = 6.4, 3.4 Hz, 1H, 3'-*H*), 3.78 (q, *J* = 3.4 Hz, 1H, 4'-*H*), 3.62 (dd, *J* = 11.8, 3.4 Hz, 1H, 5'-*H*), 3.55 (dd, *J* = 11.8, 3.4 Hz, 1H, 5'-*H*), 2.13 (ddd, *J* = 13.1, 6.4, 3.4 Hz, 1H, 2'-*H*), 1.99 (dt, *J* = 13.1, 6.4 Hz, 1H, 2'-*H*). **^{13}C NMR** (126 MHz, $(CD_3)_2SO$) δ 164.1 (*C*O), 154.3 (Ar*C*), 147.7 (Ar*C*), 87.8 (4'-*C*), 85.7 (1'-*C*), 70.4 (3'-*C*), 61.3 (5'-*C*), 56.6 (*C*NH_2), 41.3 (2'-*C*).

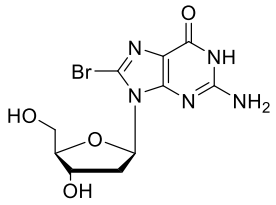
5-(trimethylsilyl)ethynyl-2'-deoxycytidine 30[194]



To a stirred solution of **29** (1.059 g, 3.0 mmol) in deoxygenated DMF (22 mL) was added bis(triphenylphosphine) palladium (II) chloride (105 mg, 0.15 mmol, 5 mol%) and copper (I) iodide (57 mg, 0.3 mmol, 10 mol%) and the yellow solution was allowed to stir under Ar for 10 minutes. The solution was treated with trimethylsilyl acetylene (2.1 mL, 15.0 mmol) and stirred for 10 minutes, then treated with triethylamine (6 mL) and allowed to stir at rt overnight. The solvents were removed *in vacuo* and purified *via* flash chromatography eluting with CH_2Cl_2/MeOH (9:1) to yield the title compound as an off-white foam (0.596 g, 62%).

MP 162–167 °C. **^{1}H NMR** (500 MHz, CD_3OD) δ 8.37 (s, 1H, Ar*H*), 6.20 (t, *J* = 6.5 Hz, 1H, 1'-*H*), 4.39 (dt, *J* = 6.5, 3.6 Hz, 1H, 3'-*H*), 3.96 (q, *J* = 3.6 Hz, 1H, 4'-*H*), 3.83 (dd, *J* = 12.0, 3.6 Hz, 1H, 5'-*H*), 3.75 (dd, *J* = 12.0, 3.6 Hz, 1H, 5'-*H*), 2.38 (ddd, *J* = 13.5, 6.5, 3.6 Hz, 1H, 2'-*H*), 2.15 (dt, *J* = 13.5, 6.5 Hz, 1H, 2'-*H*), 0.24 (s, 9H, 3 x C*H*$_3$). **^{13}C NMR** (126 MHz, CD_3OD) δ 164.6 (*C*O), 155.2 (Ar*C*), 145.0 (Ar*C*), 100.5 (*C*), 95.1 (*C*), 87.7 (4'-*C*), 86.5 (1'-*C*), 70.3 (3'-*C*), 61.0 (5'-*C*), 40.9 (2'-*C*), -1.4 (3 x *C*H$_3$).

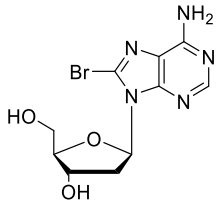
8-bromo-2'-deoxyguanosine 33[149]



To a stirred solution of 2'-deoxyguanosine (2.85 g, 10 mmol) in 4:1 MeCN/H_2O (140 mL) was added *N*-bromosuccinimide (2.65 g, 15 mmol) and the reaction was allowed to stir at rt for 45 minutes. The solution was evaporated to dryness, suspended in acetone and allowed to stir for 2 hours. The solution was cooled to -20 °C for 48 hours after which time the precipitate was collected to yield the title compound as an off-white powder (3.14 g, 86%).

MP 201–205 °C (lit. 209–213 °C)[149] **^{1}H NMR** (500 MHz, $(CD_3)_2SO$) δ 10.79 (s, 1H), 6.48 (s, 2H, N*H*$_2$), 6.16 (dd, *J* = 7.9, 6.7 Hz, 1H, 1'-*H*), 5.24 (d, *J* = 4.3 Hz, 1H), 4.85 (t, *J* = 5.9 Hz, 1H, 4'-*H*), 4.40 (dq, *J* = 5.9, 3.0 Hz, 1H, 3'-*H*), 3.62 (dt, *J* = 10.9, 5.9 Hz, 1H, 5'-*H*), 3.50 (dt, *J* = 10.9, 5.9 Hz, 1H, 5'-*H*), 3.16 (ddd, *J* = 13.2, 7.9, 6.7 Hz, 1H, 2'-*H*), 2.10 (ddd, *J* = 13.2, 6.7, 3.0 Hz, 1H, 2'-*H*). **^{13}C NMR** (126 MHz, $(CD_3)_2SO$) δ 155.8 (Ar*C*), 153.7 (Ar*C*), 152.4 (Ar*C*), 120.9 (Ar*C*), 117.9 (Ar*C*), 88.3 (4'-*C*), 85.5 (1'-*C*), 71.4 (3'-*C*), 62.5 (5'-*C*), 36.9 (2'-*C*).

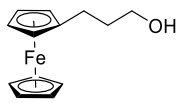
8-bromo-2'-deoxyadenosine 34[195]



To a stirred solution of 2'-deoxyadenosine (3.43 g, 22.3 mmol) in 1M acetic acid buffer (pH 4.0) (450 mL) was added Br_2 (1.3 mL, 44.6 mmol) and the solution was allowed to stir at rt for 4 hrs in the absence of light. The solution was quenched by the addition of $NaHSO_{3(sat.)(aq.)}$ (200 mL) and extracted with $CHCl_3$ (3 x 200 mL). The combined organics were concentrated and the residual AcOH was azeotroped with toluene (2 x 100 mL) to yield the compound as a white power (3.84 g, 52%).

MP 240–243 °C (lit. 242–245 °C (decomp.))[196] **^{1}H NMR** (300 MHz, $(CD_3)_2SO$) δ 8.10 (s, 1H, Ar*H*), 6.30 (dd, *J* = 8.1, 6.4 Hz, 1H, 1'-*H*), 4.48 (dt, *J* = 5.9, 2.6 Hz, 1H, 3'-*H*), 3.90 (td, *J* = 4.4, 2.6 Hz, 1H, 4'-*H*), 3.66 (dd, *J* = 12.0, 4.4 Hz, 1H, 5'-*H*), 3.49 (dd, *J* = 12.0, 4.4 Hz, 1H, 5'-*H*), 3.21 (ddd, *J* = 13.2, 8.1, 5.9 Hz, 1H, 2'-*H*), 2.19 (ddd, *J* = 13.2, 6.4, 2.6 Hz, 1H, 2'-*H*). **^{13}C NMR** (75 MHz, $(CD_3)_2SO$) δ 155.3 (Ar*C*), 152.6 (Ar*C*), 150.2 (Ar*C*), 126.9 (Ar*C*), 119.9 (ArC), 88.7 (4'-*C*), 86.8 (1'-*C*), 71.5 (3'-*C*), 62.4 (5'-*C*), 37.4 (2'-*C*).

3-ferrocenylpropan-1-ol 38[110]



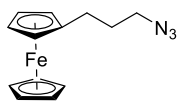
To a suspension of sodium hydride (60% dispersion in mineral oil) (1.56 g, 39 mmol) in anhydrous THF (60 mL) at 0 °C was added triethyl phosphonoacetate (7.7 mL, 39 mmol) dropwise *via* syringe and the solution was allowed to stir at rt for 30 minutes. The slurry was cooled to 0 °C and treated with a solution of ferrocenecarboxaldehyde (6.42 g, 30 mmol) in anhydrous THF (60 mL) dropwise. Once addition was complete the solution was allowed to stir at rt for 60 minutes. The solution was then poured into $NH_4Cl_{(sat.)}$ (60 mL) and the aqueous layer was extracted with EtOAc (3 x 100 mL). The combined organics were washed with $brine_{(sat.)}$ (50 mL), dried over $MgSO_4$, filtered, and concentrated *in vacuo* to give a red solid.

The red solid was dissolved in methanol (120 mL) and cooled to 0 °C, and treated with 10% palladium on carbon (1.57 g, 5 mol%) and ammonium formate (7.50 g 120 mmol). Once addition was complete the suspension was warmed to rt. The suspension was then stirred for 4 hours, then filtered through celite, washing the solids with methanol until the washings ran clear. The orange solution was concentrated *in vacuo*, and partitioned between EtOAc (100 mL) and $NaHCO_{3\ (sat.)}$ (50 mL). The aqueous layer was extracted with EtOAc (3 x 50 mL) and the combined organics were washed with $brine_{(sat.)}$ (50 mL), dried over $MgSO_4$, filtered, and concentrated *in vacuo* to yield a yellow solid.

To a stirred suspension of $LiAlH_4$ (3.40 g, 90 mmol) in anhydrous Et_2O (150 mL) at 0 °C was added a solution of the yellow solid in anhydrous Et_2O (150 mL) dropwise *via* a pressure equalising funnel. Once addition was complete, the solution was warmed to rt and allowed to stir for 60 minutes. The solution was cooled to 0 °C and quenched sequentially with H_2O (3.4 mL), 15% $NaOH_{(aq.)}$ (3.4 mL), and H_2O (10.2 mL). The slurry was stirred at rt for 10 minutes, then filtered through celite, washing with Et_2O until the washings ran clear. The yellow solution was dried over $MgSO_4$, filtered, and concentrated *in vacuo*, yielding a yellow oil which was purified *via* silica gel chromatography, eluting with petrol/EtOAc (8:2) to yield the title compound as a yellow oil (6.10 g, 83%).

^{1}H NMR (300 MHz, C_6D_6) δ 4.01 (s, 5H, Cp*H*), 3.98–3.94 (m, 4H, Cp*H*), 3.41 (t, *J* =6.4 Hz, 2H, CH_2), 2.29 (dd, *J* = 8.6, 6.4 Hz, 2H, CH_2), 1.64–1.59 (m, 2H, CH_2). **^{13}C NMR** (126 MHz, C_6D_6) δ 88.9 (Cp*C*), 69.2 (Cp*C*H), 68.6 (Cp*C*H), 67.4 (Cp*C*H), 62.2 (*C*H_2), 34.5 (*C*H_2), 26.1 (*C*H_2).

1-azido-3-ferrocenylpropane 39

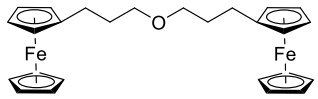


To a stirred solution of **38** (1.22 g, 5 mmol) and tetrabromomethane (1.82 g, 5.5 mmol) in anhydrous CH_2Cl_2 (12.5 mL) at 0 °C was added triphenylphosphine (1.44 g, 5.5 mmol) portionwise over a 10-minute period. The reaction was allowed to warm to rt and stirred for 30 minutes. The solution was concentrated *in vacuo* to give a thick oil, which was suspended in hexane (50 mL) and stirred vigorously for 30 minutes. The solution was filtered, and the slurry was suspended further with hexane (50 mL) and filtered. The combined organics were concentrated *in vacuo* and filtered through silica, washing with EtOAc until the washings ran clear. The organics were concentrated *in vacuo* to yield a brown oil.

The oil was dissolved in acetone (15 mL) and H_2O (5 mL) and treated with sodium azide (0.97 g, 15 mmol) and the solution was heated at 70 °C for 18 hours. The reaction was cooled to rt and diluted with EtOAc (25 mL) and water (10 mL). The aqueous layer was extracted with EtOAc (3 x 20 mL) and the combined organics were washed with $brine_{(sat.)}$ (20 mL), dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The yellow oil was purified *via* silica gel chromatography, eluting with petrol/EtOAc (1:1) to yield the title compound as a yellow oil (1.30 g, 93%).

IR ν (cm^{-1}) 3089, 2933, 2857, 2094. **^{1}H NMR** (300 MHz, C_6D_6) δ 3.97 (s, 5H, Cp*H*), 3.92 (t, *J* = 1.8 Hz, 2H, Cp*H*), 3.85 (t, *J* = 1.8 Hz, 2H, Cp*H*), 2.72 (t, *J* = 6.7 Hz, 2H, C*H*$_2$), 2.09 (dd, *J* = 8.4, 6.7 Hz, 2H, C*H*$_2$), 1.47–1.33 (m, 2H, C*H*$_2$). **^{13}C NMR** (75 MHz, C_6D_6) δ 88.1 (Cp*C*), 69.3 (Cp*C*H), 68.7 (Cp*C*H), 68.1 (Cp*C*H), 51.2 (*C*H$_2$), 30.8 (*C*H$_2$), 27.1 (*C*H$_2$). **HRMS** (ESI) calculated for $C_{13}H_{15}FeN_3$ $[M]^{+\cdot}$ *m/z* 269.0615, found 269.0595.

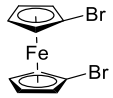
2,2'-(oxybis(propane-3,1-diyl))diferrocene 41



To a stirred solution of **38** (1.22 g, 5 mmol) in anhydrous THF (50 mL) was added tosyl chloride (0.95 g, 5 mmol) and triethyl amine (0.80 mL, 5.5 mmol) and the reaction heated to 60 °C overnight. The reaction poured into $NaHCO_3$ (50 mL) and the aqueous layer was extracted with EtOAc (3 x 50 mL). The combined organics were washed with brine $_{(sat.)}$ (50 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude oil was purified *via* silica gel chromatography, eluting with neat petrol to yield the title compound as an undesired by-product (84 mg, 30%).

IR ν (cm^{-1}) 3076, 2967, 2869. **^{1}H NMR** (250 MHz, C_6D_6) δ 3.97 (s, 10H, Cp*H*), 3.92 (t, *J* = 1.8 Hz, 4H, Cp*H*), 3.86 (t, *J* = 1.8 Hz, 4H, Cp*H*), 3.14 (t, *J* = 6.4 Hz, 4H, Cp*H*), 2.29–2.18 (m, 4H, Cp*H*), 1.71–1.56 (m, 4H, Cp*H*). **^{13}C NMR** (63 MHz, C_6D_6) δ 87.5 (Cp*C*), 68.9 (Cp*C*H), 68.4 (Cp*C*H), 67.7 (Cp*C*H), 44.5 (*C*H_2), 34.2 (*C*H_2), 26.9 (*C*H_2). **HRMS** (ESI) calculated for $C_{26}H_{31}Fe_2O$ $[M+H]^+$: *m/z* 471.1068 found 471.1071.

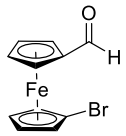
1,1'-dibromoferrocene 44[150]



To an oven-dried three-neck flask was added ferrocene (18.60 g, 0.1 mol) and *N,N,N',N'*-tetramethylethylene diamine (35 mL, 0.233 mol) and were suspended in anhydrous hexane (100 mL). The suspension was cooled to 0 °C and *n*-butyllithium (2.5 M in hexanes) (85 mL, 0.212 mol) was added to a pressure equalising funnel, then added dropwise to the ferrocene solution. The suspension was warmed to rt and allowed to stir overnight. After this time, the orange precipitate was allowed to settle, and the hexane was removed *via* cannula filtration. The bright orange precipitate was suspended in anhydrous Et_2O (300 mL) and cooled to -78 °C. To the orange suspension was added 1,1,2,2-tetrabromoethane (25 mL, 0.215 mol) in Et_2O (150 mL) dropwise over 4 hours, maintaining -78 °C throughout, after which time the solution was allowed to warm to rt overnight. The solution was allowed to settle to give a biphasic mixture, and the top orange layer was decanted. The solution was washed with H_2O (100 mL) and the aqueous layer was extracted with Et_2O (3 x 100 mL). The solution was concentrated, and the orange oil was dissolved in hexane (100 mL) and washed with 2M $FeCl_{3(aq.)}$ (2 x 200 mL) and the combined organics were washed with H_2O (3 x 100 mL). The organics were dried over $MgSO_4$, filtered, and concentrated. The orange oil was dissolved in the minimum amount of hot MeOH (~15 mL) and allowed to recrystallise at –20 °C overnight. The solids were filtered, washed with cold MeOH and dried under reduced pressure to yield the title compound as an orange solid (11.09 g, 32%).

MP 53–54 °C. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.16 (t, *J* = 1.9 Hz, 4H, Cp*H*), 3.71 (t, *J* = 1.9 Hz, 4H, Cp*H*). **^{13}C NMR** (126 MHz, C_6D_6) δ 79.4 (Cp*C*), 71.6 (Cp*C*H), 68.3 (Cp*C*H).

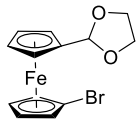
1'-bromoferrocenecarboxaldehyde 45[197]



To an oven-dried Schlenk tube equipped with magnetic stirrer bar was added **44** (6.88 g, 20 mmol). The flask was sealed, then evacuated and back-filled with argon four times, and left over an argon atmosphere. The solid was dissolved in anhydrous THF (70 mL) and cooled to –78 °C. To the orange solution was added *n*-butyllithium (2.5 M in hexanes) (8 mL, 20 mmol) dropwise and after addition was complete, the solution was allowed to stir at –78 °C for 1 hour. The orange precipitate was then treated with dimethyl formamide (4.6 mL, 60 mmol) and the solution was allowed to warm to rt and stirred overnight. The solution was quenched with 2M $NaOH_{(aq.)}$ (40 mL) and the aqueous layer was extracted with Et_2O (3 x 100 mL). The combined organics were washed with $brine_{(sat.)}$ (100 mL), dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The oil was purified *via* flash chromatography, eluting with petrol/EtOAc (9:1) to yield the title compound as a deep red solid (5.53g 94%).

MP 52–53 °C. **^{1}H NMR** (500 MHz, C_6D_6) δ 9.79 (s, 1H, C*H*O), 4.49 (s, 2H, Cp*H*), 4.10 (s, 2H, Cp*H*), 4.05 (s, 2H, Cp*H*), 3.61 (s, 2H, Cp*H*). **^{13}C NMR** (126 MHz, C_6D_6) δ 191.8 (*C*HO), 81.7 (Cp*C*), 78.0 (Cp*C*), 75.3 (Cp*C*H), 71.7 (Cp*C*H), 71.6 (Cp*C*H), 68.6 (Cp*C*H).

(1'-bromo)ferrocenyl-1,3-dioxolane 46



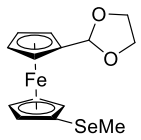
To a stirred solution of **45** (3.37 g, 10 mmol) and *p*-toluenesulfonic acid (170 mg, 1 mmol) in anhydrous benzene (60 mL) was added ethylene glycol (5.57 mL, 100 mmol). The flask was fitted with a Dean-Stark trap and condenser, and allowed to stir under reflux for 18 hours. The solution was quenched with K_2CO_2 (3 g) and stirred for 10 minutes, filtered through celite and concentrated under reduced pressure. The crude dark solid was purified *via* silica gel chromatography, eluting with petrol/EtOAc (95:5) to yield an orange solid (2.99 g, 78%).

MP 58–60 °C **IR** ν (cm^{-1}) 3097, 2918, 2800. **^{1}H NMR** (500 MHz, C_6D_6) δ 5.69 (s, 1H, FcC*H*O_2), 4.35 (t, *J* = 2.0 Hz, 2H, Cp*H*), 4.29 (t, *J* = 2.0 Hz, 2H, Cp*H*), 4.00 (t, *J* = 2.0 Hz, 2H, Cp*H*), 3.90 (t, *J* = 2.0 Hz, 2H, Cp*H*), 3.64–3.55 (m, 2H, CH_2), 3.51–3.41 (m, 2H, CH_2). **^{13}C NMR** (126 MHz, C_6D_6) δ 102.4 (Fc*C*HO_2), 87.0 (Cp*C*), 78.4 (Cp*C*), 71.4 (Cp*C*H), 71.4 (Cp*C*H), 70.0 (Cp*C*H), 68.4 (Cp*C*H), 65.2 (*C*H_2). **HRMS** (ESI) calculated for $C_{13}H_{13}BrFeO_2$ $[M+H]^+$: *m/z* 336.9521, found 336.9526.

General procedure A for the electrophilic addition to bromoferrocene

An oven dried Schlenk flask was charged with (1'-bromo)ferrocenyl-1,3-dioxolane **46** (1 equiv.) under argon. The ferrocene was dissolved in anhydrous THF (0.2 M), cooled to -78 °C, and treated with a solution of *n*-butyllithium (1 equiv.) dropwise. The resulting yellow precipitate was allowed to stir at -78 °C for 30 minutes, after which time it was treated with the desired electrophile (1 equiv.) at -78 °C and allowed to warm to rt overnight. The solution was poured into 2M $NaOH_{(aq.)}$ (10 mL/mmol) and extracted with Et_2O (3 x 50 mL/mmol). The combined organics were washed with $brine_{(sat.)}$, dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude oil was purified *via* silica gel chromatography to yield the desired compounds.

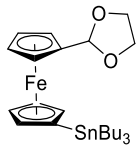
(1'-trimethylselanyl)ferrocenyl-1,3-dioxolane 47



General procedure A was followed using **46** (1.65 g, 5 mmol), *n*-butyllithium (2.5 M in hexanes) (2.0 mL), and dimethyl diselenide (475 µL, 5 mmol), eluting with petrol/EtOAc (9:1) to yield the title compound as a yellow oil (1.54 g, 88%).

IR ν (cm^{-1}) 3083, 2908, 2815. **^{1}H NMR** (500 MHz, C_6D_6) δ 5.76 (s, 1H, FcC*H*O_2), 4.36–4.31 (m, 4H, Cp*H*), 4.11 (t, *J* = 1.9 Hz, 2H, Cp*H*), 4.04 (t, *J* = 1.9 Hz, 2H (Cp*H*), 3.66–3.56 (m, 2H, CH_2), 3.53–3.43 (m, 2H, CH_2), 1.90 (s, 3H, 2 x CH_3). **^{13}C NMR** (126 MHz, C_6D_6) δ 102.4 (Fc*C*HO_2), 85.9 (Cp*C*), 74.3 (Cp*C*H), 70.1 (Cp*C*H), 69.7 (Cp*C*H), 68.8 (Cp*C*), 68.4 (Cp*C*H), 64.8 (*C*H_2), 8.9 (Se*C*H_3). **HRMS** (ESI) calculated for $C_{14}H_{17}FeO_2Se$ $[M+H]^+$: *m/z* 352.9738, found 352.9749.

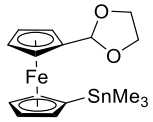
(1'-tributylstannyl)ferrocenyl-1,3-dioxolane 48



General procedure A was followed using **46** (1.65 g, 5 mmol), *n*-butyllithium (2.5 M in hexanes) (2.0 mL), and tributyltin chloride (1.35 mL, 5 mmol), eluting with petrol/EtOAc (9:1) to yield the title compound as a yellow oil (2.32 g, 64%).

IR ν (cm^{-1}) 3094, 2921 2820. **^{1}H NMR** (500 MHz, C_6D_6) δ 5.79 (s, 1H, $FcCHO_2$), 4.46–4.40 (m, 2H, Cp*H*), 4.40 (t, *J* = 1.9 Hz, 2H, Cp*H*), 4.15–4.12 (m, 2H, Cp*H*), 4.09 (t, *J* = 1.9 Hz, 2H, Cp*H*), 3.68–3.58 (m, 2H, CH_2), 3.54–3.44 (m, 2H, CH_2), 1.76–1.52 (m, 6H, 3 x CH_2), 1.42–1.30 (m, 1H, 3 x CH_2), 1.19–1.02 (m, 6H, 3 x CH_2), 0.95 (t, *J* = 7.3 Hz, 9H, 3 x CH_3). **^{13}C NMR** (126 MHz, C_6D_6) δ 103.5 ($FcCHO_2$), 85.6 (Cp*C*), 75.8 (Cp*C*H), 72.2 (Cp*C*H), 69.7 (Cp*C*), 68.8 (Cp*C*H), 67.6 (Cp*C*H), 65.4 (CH_2), 29.9 (CH_2), 28.0 (CH_2), 14.2 (CH_2), 10.7 (CH_2). **HRMS** (ESI) calculated for $C_{25}H_{40}FeO_2SnNa$ $[M+Na]^+$: *m/z* 571.1292, found 571.1301.

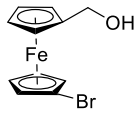
(1'-trimethylstannyl)ferrocenyl-1,3-dioxolane 96



General procedure A was followed using **46** (1.65 g, 5 mmol), *n*-butyllithium (2.5 M in hexanes) (2.0 mL), and trimethyltin chloride (995 μL, 5 mmol), eluting with petrol/EtOAc (9:1) to yield the title compound as a yellow oil (1.51 g, 72%).

IR ν (cm^{-1}) 3086, 2899, 2807. **^{1}H NMR** (500 MHz, C_6D_6) δ 5.75 (s, 1H, $FcCHO_2$), 4.42–4.37 (m, 2H, Cp*H*), 4.35–4.33 (m, 2H, Cp*H*), 4.12–4.05 (m, 2H, Cp*H*), 4.01 (d, *J* = 1.8 Hz, 2H, Cp*H*), 3.67–3.56 (m, 2H, CH_2), 3.53–3.41 (m, 2H, CH_2), 0.26 (s, 9H, 3 x CH_3). **^{13}C NMR** (126 MHz, C_6D_6) δ 103.2 ($FcCHO_2$), 85.5 (Cp*C*), 75.3 (Cp*C*H), 72.0 (Cp*C*H), 69.6 (Cp*C*), 68.5 (Cp*C*H), 67.4 (Cp*C*H), 65.2 (CH_2), -8.7 ($SnCH_3$). **HRMS** (ESI) calculated for $C_{16}H_{22}FeNaO_2S$ $[M+Na]^+$: *m/z* 444.9883, found 444.9898.

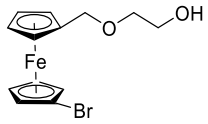
(1'-bromoferrocenyl)methanol 49



To a stirred slurry of $LiAlH_4$ (277 mg, 7.5 mmol) in anhydrous THF (5 mL) at 0 °C was added a solution of **45** (1.47 g, 5.0 mmol) in anhydrous THF (5 mL) dropwise, and once addition was complete the suspension was allowed to stir at rt for 1 hour. The yellow slurry was cooled to 0 °C and quenched sequentially with water (277 µL), 15% $NaOH_{(aq.)}$ (277 µL), then water (831 µL) and stirred at rt for 15 minutes. The suspension was filtered through celite, dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude oil was purified *via* flash chromatography, eluting with petrol/EtOAc (8:2) to yield the title compound as a yellow oil (1.28 g, 86%).

IR ν (cm^{-1}) 3310, 3101, 2897. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.23 (s, 2H, C*H*$_2$), 4.17 (t, *J* = 1.9 Hz, 2H, Cp*H*), 4.05 (t, *J* = 1.9 Hz, 2H, Cp*H*), 3.94 (t, *J* = 1.9 Hz, 2H, Cp*H*), 3.70 (t, *J* = 1.9 Hz, 2H, Cp*H*). **^{13}C NMR** (126 MHz, C_6D_6) δ 90.4 (Cp*C*), 78.3 (Cp*C*H), 70.8 (Cp*C*H), 70.4 (Cp*C*H), 68.6 (Cp*C*), 67.7 (Cp*C*H), 60.1 (C*H*$_2$OH). **HRMS** (ESI) calculated for $C_{11}H_{12}BrFeO$ $[M+H]^+$: *m/z* 294.9415, found 294.9423.

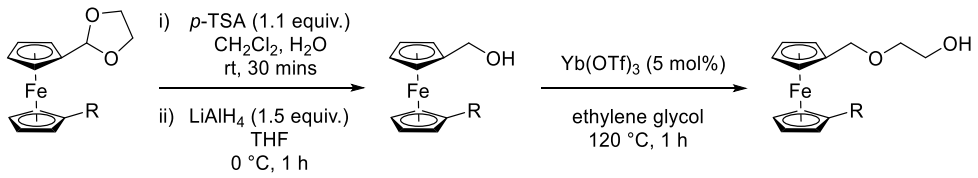
2-((1'-bromo)ferrocenyl)methyloxyethan-1-ol 50



To a stirred solution of **49** (710 mg, 2.4 mmol) in ethylene glycol (24 mL) was added ytterbium (III) triflate (74 mg, 0.12 mmol), and the solution was heated to 120 °C for 20 minutes. The reaction was cooled, diluted with H_2O (200 mL), and extracted with EtOAc (3 x 100 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude orange oil was purified *via* flash chromatography, eluting with petrol/EtOAc (8:2) to yield the title compound as an orange oil (703 mg, 69%).

IR ν (cm^{-1}) 3287, 3057, 2953, 2889. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.30 (s, 2H, FcC*H*$_2$), 4.25 (s, 2H, Cp), 4.18 (s, 2H, Cp), 4.07 (s, 2H, Cp), 3.83 (s, 2H, Cp), 3.71–3.59 (m, 2H, C*H*$_2$), 3.47–3.35 (m, 2H, C*H*$_2$), 2.76 (s, 1H, O*H*). **^{13}C NMR** (126 MHz, C_6D_6) δ 85.5 (Cp*C*), 78.5 (Cp*C*), 72.0 (Cp*C*H), 71.8 (CH_2), 71.2 (Cp*C*H), 71.0 (Cp*C*H), 68.8 (CH_2), 67.9 (Cp*C*H), 61.9 (CH_2). **HRMS** (ESI) calculated for $C_{13}H_{15}BrFeO_2Na$ $[M+Na]^+$: *m/*z 360.9497, found 360.9500.

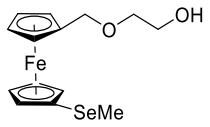
General procedure B



To a stirred solution of acetal (1 equiv.) in dichloromethane (3.3 mL/mmol) and water (1.6 mL/mmol) under argon was added *p*-toluenesulfonic acid (1.1 equiv.) and the mixture was allowed to stir at rt for 2 hours. The solution was quenched with $NaHCO_{3(sat.)(aq.)}$ (10 mL/mmol) and the aqueous layers were extracted with EtOAc (3 x 25 mL/mmol). The combined organics were washed with brine, dried over $MgSO_4$, filtered, and concentrated *in vacuo*.

To a stirred slurry of $LiAlH_4$ (1.5 equiv.) in anhydrous THF (1 mL/mmol) at 0 °C was added a solution of the crude oil in anhydrous THF (1 mL/mmol) dropwise, and once addition was complete the suspension was allowed to stir at rt for 1 hour. The reaction was cooled to 0 °C and quenched sequentially with H_2O (1 µL per mg $LiAlH_4$), 15% $NaOH_{(aq.)}$ (1 µL per mg $LiAlH_4$), and H_2O (3 µL per mg $LiAlH_4$). The suspension was filtered through celite, dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude oil was dissolved in ethylene glycol (0.1 M), treated with ytterbium (III) triflate (5 mol%), and heated to 120 °C for 20 minutes. The reaction was cooled and poured into H_2O (100 mL/mmol) and extracted with EtOAc (3 x 50 mL/mmol). The combined organics were washed with $brine_{(sat.)}$, dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude oil was purified *via* flash chromatography to yield the desired compound.

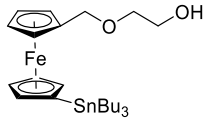
2-((1'-methylseleno)ferrocenyl)methyloxyethan-1-ol 53



General procedure B was followed using **47** (1.40 g, 4 mmol), *p*-toluenesulfonic acid (756 mg, 4.4 mmol) in CH_2Cl_2 (15 mL) and water (7 mL), then $LiAlH_4$ (222 mg, 6 mmol) in THF (6 mL), and finally ytterbium (III) triflate (124mg, 0.2 mmol) in ethylene glycol (40 mL), eluting with petrol/EtOAc (8:2) to yield the title compound as a yellow oil (708 mg, 50%).

IR ν (cm^{-1}) 3219, 3048, 2931, 2876. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.21 (t, *J* = 1.8 Hz, 2H, Cp*H*), 4.19 (s, 2H, FcC*H*$_2$), 4.12 (t, *J* = 1.9 Hz, 2H, Cp*H*), 4.01 (t, *J* = 1.8 Hz, 2H, Cp*H*), 3.93 (t, *J* = 1.9 Hz, 2H, Cp*H*), 3.60 (dd, *J* = 9.5, 6.0 Hz, 2H, C*H*$_2$), 3.37–3.33 (m, 2H, C*H*$_2$), 2.67 (t, *J* = 6.0 Hz, 1H, O*H*), 1.88 (s, 3H, SeC*H*$_3$). **^{13}C NMR** (126 MHz, C_6D_6) δ 85.2 (Cp*C*), 74.8 (Cp*C*H), 71.9 (*C*H$_2$), 70.8 (Cp*C*), 70.1 (Cp*C*H), 69.9 (Cp*C*H), 69.1 (*C*H$_2$), 68.8 (Cp*C*), 61.9 (*C*H$_2$), 9.62 (Se*C*H$_3$). **HRMS** (ESI) calculated for $C_{14}H_{18}FeO_2SeNa$ [M+Na]$^+$: *m/z* 376.9714, found 376.9728.

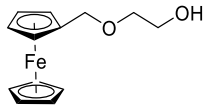
2-((1'-tributylstannyl)ferrocenyl)methyloxyethan-1-ol 54



General procedure B was followed using **48** (2.178 g, 3 mmol), *p*-toluenesulfonic acid (567 mg, 3.3 mmol) in CH_2Cl_2 (10 mL) and water (5 mL), then $LiAlH_4$ (166 mg, 4.5 mmol) in THF (6 mL), and finally ytterbium (III) triflate (93 mg, 0.15 mmol) in ethylene glycol (30 mL), eluting with petrol/EtOAc (8:2) to yield the title compound as a yellow oil (474 mg, 29%).

IR ν (cm^{-1}) 3310, 3058, 2947, 2913, 2847. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.22 (s, 2H, FcC*H*$_2$), 4.21 (t, *J* = 1.6 Hz, 2H, Cp*H*), 4.16 (t, *J* = 1.8 Hz, 2H, Cp*H*), 4.06 (t, *J* = 1.8 Hz, 2H, Cp*H*), 3.99 (t, *J* = 1.6 Hz, 2H, Cp*H*), 3.56 (q, *J* = 5.6 Hz, 2H, C*H*$_2$), 3.38–3.34 (m, 2H, C*H*$_2$), 2.35 (t, *J* = 5.6 Hz, 1H, O*H*), 1.71–1.57 (m, 6H, 3 x C*H*$_2$), 1.40 (h, *J* = 7.4 Hz, 6H, 3 x C*H*$_2$), 1.12–1.04 (m, 6H, 3 x C*H*$_2$), 0.95 (t, *J* = 7.4 Hz, 9H, 3 x C*H*$_3$). **^{13}C NMR** (126 MHz, C_6D_6) δ 83.7 (Cp*C*), 75.2 (Cp*C*H), 71.6 (*C*H$_2$), 71.5 (Cp*C*H), 69.7 (Cp*C*H), 69.1 (Cp*C*), 68.8 (Cp*C*H), 61.9 (*C*H$_2$), 29.7 (*C*H$_2$), 27.8 (*C*H$_2$), 14.0 (*C*H$_2$), 10.6 (*C*H$_3$). **HRMS** (ESI) calculated for $C_{25}H_{42}FeO_2SnNa$ [M+Na]$^+$: *m/z* 573.1448, found 573.1454.

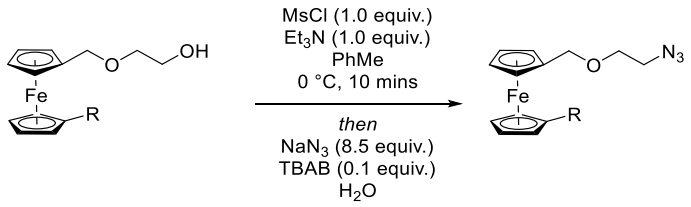
2-(ferrocenyl)methyloxyethan-1-ol 55



55 was isolated as a by-product of the formation of **54** due to protodestannylation to yield the title compound as a yellow oil (117 mg, 15%).

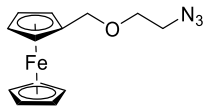
IR ν (cm^{-1}) 3289, 3063, 2894. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.14 (s, 2H, FcC*H*$_2$), 4.11 (t, *J* = 1.8 Hz, 2H, Cp*H*), 3.97 (s, 5H, Cp*H*), 3.97 (t, *J* = 1.8 Hz, 2H, Cp*H*), 3.59 (t, *J* = 4.8 Hz, 2H, C*H*$_2$), 3.37–3.33 (m, 2H, C*H*$_2$), 2.77 (s, 1H, O*H*). **^{13}C NMR** (126 MHz, C_6D_6) δ 83.9 (Cp*C*), 71.6 (*C*H$_2$), 69.7 (Cp*C*H), 69.5 (*C*H$_2$), 68.8 (Cp*C*H), 68.7 (Cp*C*H), 61.9 (*C*H$_2$). **HRMS** (ESI) calculated for $C_{13}H_{16}FeO_2Na$ $[M+Na]^+$: *m/z* 283.0392, found 283.0422.

General procedure C for the azidation of ferrocenyl alcohols



To a stirred solution of ferrocenyl alcohols (1 equiv.) in toluene (4 mL/mmol) at 0 °C was added methanesulfonyl chloride (1 equiv.) in one portion and allowed to stir for 5 minutes. Triethylamine (1 equiv.) was added dropwise and allowed to stir for 5 minutes. To this solution was added water (2 mL/mmol), tetrabutylammonium bromide (0.1 equiv.) and sodium azide (8.5 equiv.). The flask was fitted with a reflux condenser and allowed to stir at 80 °C overnight. The reaction was cooled and the aqueous layer was extracted with EtOAc (3 x 25 mL/mmol). The combined organics were washed with brine$_{(sat.)}$ (25 mL/mmol), dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude oil was purified *via* flash chromatography to yield the desired compound.

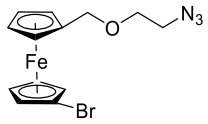
((2-azido ethoxy)methyl)ferrocene 56



General procedure C was followed using **55** (117 mg, 0.45 mmol) eluting with petrol/EtOAc (9:1) to yield the title compound as an orange oil (80 mg, 62%).

IR ν (cm^{-1}) 3093, 2925, 2854, 2091. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.08 (s, 2H, Cp*H*), 4.06 (s, 2H, FcC*H*$_2$), 3.98 (s, 5H, Cp*H*), 3.96 (s, 2H, Cp*H*), 3.17 (t, *J* = 5.0 Hz, 2H, C*H*$_2$), 2.78 (t, *J* = 5.0 Hz, 2H, C*H*$_2$). **^{13}C NMR** (126 MHz, C_6D_6) δ 83.8 (Cp*C*), 69.5 (*C*H$_2$), 69.4 (Cp*C*H), 68.8 (Cp*C*H), 68.7 (*C*H$_2$), 68.7 (Cp*C*H), 50.8 (*C*H$_2$). **HRMS** (ESI) calculated for $C_{13}H_{15}FeN_3ONa$ [M+Na]$^+$: *m/z* 308.0457, found 308.0462.

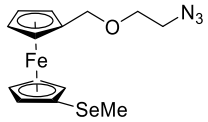
1'-bromo-1-((2-azidoethoxy)methyl)ferrocene 57



General procedure C was followed using **49** (560 mg, 1.6 mmol), eluting with petrol/EtOAc (9:1) to yield the title compound as an orange oil (230 mg, 39%).

IR ν (cm^{-1}) 3098, 2858, 2096. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.29–3.85 (m, 8H, C*H*$_2$, 3 x Cp*H*), 3.85–3.62 (m, 2H, Cp*H*), 3.22–2.99 (m, 2H, C*H*$_2$), 2.87–2.62 (m, 2H, C*H*$_2$). **^{13}C NMR** (126 MHz, C_6D_6) δ 86.2 (Cp*C*), 78.5 (Cp*C*), 71.9 (Cp*C*H), 71.0 (Cp*C*H), 71.0 (Cp*C*H), 70.0 (*C*H$_2$), 68.8 (*C*H$_2$), 68.0 (Cp*C*H), 50.8 (*C*H$_2$). **HRMS** (ESI) calculated for $C_{13}H_{14}BrFeN_3ONa$ [M+Na]$^+$: *m/z* 385.9562, found 385.9567.

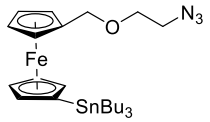
1'-methylseleno-1-((2-azidoethoxy)methyl)ferrocene 58



General procedure C was followed using **53** (353 mg, 1.0 mmol), eluting with petrol/EtOAc (9:1) to yield the title compound as an orange oil (289 mg, 76%).

IR ν (cm^{-1}) 3088, 2926, 2855, 2090. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.19 (t, *J* = 1.8 Hz, 2H, Cp*H*), 4.13 (s, 2H, FcC*H*$_2$), 4.10 (t, *J* = 1.9 Hz, 2H, Cp*H*), 4.00 (t, *J* = 1.9 Hz, 2H, Cp*H*), 3.94 (t, *J* = 1.8 Hz, 2H, Cp*H*), 3.18 (t, *J* = 5.0 Hz, 2H, C*H*$_2$), 2.80 (t, *J* = 5.0 Hz, 2H, C*H*$_2$), 1.88 (s, 3H, C*H*$_3$). **^{13}C NMR** (126 MHz, C_6D_6) δ 84.7 (Cp*C*), 74.6 (Cp*C*H), 70.7 (Cp*C*H), 70.1 (Cp*C*H), 70.0 (Cp*C*), 69.0 (*C*H$_2$), 68.9 (Cp*C*), 68.8 (*C*H$_2$), 50.8 (*C*H$_2$), 9.5 (*C*H$_3$). **HRMS** (ESI) calculated for $C_{14}H_{17}FeN_3OSeNa$ [M+Na]$^+$: *m/z* 401.9778, found 401.9779.

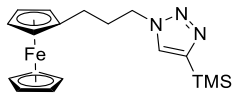
1'-tributylstannyl-1-((2-azidoethoxy)methyl)ferrocene 59



General procedure C was followed using **54** (474 mg, 0.86 mmol), eluting with petrol/EtOAc (9:1) to yield the title compound as an orange oil (280 mg, 55%).

IR ν (cm^{-1}) 3091, 2955, 2924, 2853, 2100. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.23 (t, *J* = 1.6 Hz, 2H, Cp*H*), 4.19–4.14 (m, 4H, Cp*H*, CH_2), 4.06 (t, *J* = 1.9 Hz, 2H, Cp*H*), 4.01 (t, *J* = 1.9 Hz, 2H, Cp*H*), 3.24–3.20 (m, 2H, CH_2), 2.81–2.76 (m, 2H, CH_2), 1.71–1.60 (m, 6H, 3 x CH_2), 1.45–1.36 (m, 6H, 3 x CH_2), 1.13–1.06 (m, 6H, 3 x CH_2), 0.96 (t, *J* = 7.3 Hz, 9H, 3 x CH_3). **^{13}C NMR** (126 MHz, C_6D_6) δ 83.2 (Cp*C*), 74.8 (Cp*C*H), 71.2 (Cp*C*H), 69.5–68.9 (m, Cp*C*H), 69.0 (*C*H_2), 68.4–68.2 (m, Cp*C*H), 68.3 (*C*H_2), 50.3 (*C*H_2), 29.2 (3 x *C*H_2), 27.4 (3 x *C*H_2), 13.5 (3 x *C*H_2), 10.2 (3 x *C*H_3). **HRMS** (ESI) calculated for $C_{25}H_{41}FeN_3OSnNa$ $[M+Na]^+$: *m/z* 598.1513, found 598.1545.

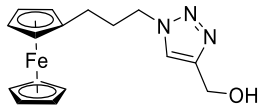
1-(3-(ferrocenyl)propyl)-4-(trimethylsilyl)-1*H*-1,2,3-triazole 60



To a stirred solution of **39** (266 mg, 1 mmol) and trimethylsilyl acetylene (0.14 mL, 1 mmol) in *t*BuOH (2.5 mL) and water (2.5 mL) was added $CuSO_4 \cdot 5H_2O$ (25 mg, 0.1 mmol) and sodium ascorbate (20 mg, 0.1 mmol) and the solution was allowed to stir for 4 hours at rt. The solution was dissolved in EtOAc (10 mL), and the aqueous layer was extracted with EtOAc (2 x 10 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The oil was purified *via* flash chromatography, eluting with n-hexane/EtOAc (4:1) to yield the title compound as a brown solid (227 mg, 62%).

IR ν (cm^{-1}) 3124, 3093, 2958, 2940, 2898. **^{1}H NMR** (300 MHz, C_6D_6) δ 7.03 (s, 1H, C*H*), 4.05–4.02 (m, 7H, Cp*H*), 4.00–3.94 (m, 4H, Cp*H*, CH_2), 2.18 (d, *J* = 7.1 Hz, 2H, CH_2), 1.82 (p, *J* = 7.1 Hz, 2H, CH_2), 0.46 (s, 9H, CH3). **^{13}C NMR** (75 MHz, C_6D_6) δ 145.8 (Ar*C*), 87.4 (Cp*C*), 68.9 (Cp*C*H), 68.4 (Cp*C*H), 67.7 (Cp*C*H), 48.7 (*C*H_2), 32.0 (*C*H_2), 26.5 (*C*H_2), –0.78 (*C*H_3). **HRMS** (ESI) calculated for $C_{18}H_{26}FeN_3Si$ $[M+H]^+$: *m/z* 368.1245, found: 368.1275.

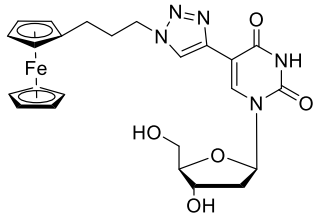
(1-(3-(thiophen-2-yl)propyl)-1*H*-1,2,3-triazol-4-yl)methanol 61



To a stirred solution of **39** (1.07 mg, 4 mmol) and propargyl alcohol (0.23 mL, 4 mmol) in tBuOH (5 mL) and H_2O (5 mL) was added $CuSO_4 \cdot 5H_2O$ (100 mg, 0.4 mmol) and sodium ascorbate (80 mg, 0.4 mmol) and allowed to stir for 18 hours at rt. The solution was dissolved in EtOAc (10 mL) and the aqueous layer was extracted with EtOAc (2 x 10 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The oil was purified *via* flash chromatography, eluting with *n*-hexane/EtOAc (4:1) to yield the title compound as a brown solid (1.23 g, 95%).

IR ν (cm^{-1}) 3233, 3123, 2981, 2931, 2851. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.52 (s, 1H, Ar*H*), 4.82 (s, 2H, C*H*$_2$), 4.35 (t, *J* = 1.7 Hz, 2H, Cp*H*), 4.18–4.03 (m, 9H, Cp*H,* C*H*$_2$), 2.61 (s, 1H, O*H*), 2.39 (t, *J* = 7.5 Hz, 2H, C*H*$_2$), 2.16–2.07 (m, 2H, CH_2). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 147.3 (Ar*C*), 121.2 (Ar*C*), 86.6 (Cp*C*), 68.3 (Cp*C*H), 67.8 (Cp*C*H), 67.2 (Cp*C*H), 56.4 (O*C*H$_2$), 49.5 (*C*H$_2$), 31.4 (*C*H$_2$), 26.2 (*C*H$_2$). **HRMS** (ESI) calculated for $C_{16}H_{19}FeN_3ONa$ $[M+Na]^+$: *m/z* 348.0775, found: 348.0777.

5-((3-ferrocenylpropyl)-1*H*-1,2,3-triazol-4-yl)-2'-deoxyuridine 62



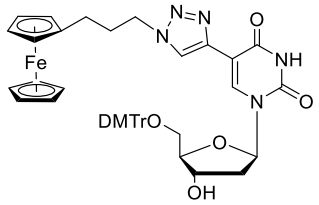
To a stirred solution of **28** (1.14 g, 3.5 mmol) in MeOH (12 mL) was added K_2CO_3 (1.93 g, 14 mmol) and the solution was allowed to stir at room temperature for 30 minutes. The solution was filtered and concentrated and the colourless gum was dissolved in EtOAc (7 mL) and H_2O (7 mL) with **39** (0.94 g, 3.5 mmol). The yellow solution was treated with $CuSO_4 \cdot 5H_2O$ (87 mg, 0.35 mmol) and sodium ascorbate (138 mg, 0.7 mmol) and the solution was heated to 70 °C for 3 hours. The solvent was removed and the yellow residue was filtered through silica eluting with $CDCl_3$/MeOH (9:1). The filtrate was dried with $MgSO_4$ and concentrated to give a yellow gum. The gum was recrystallised from hot MeOH to yield a yellow powder (867 mg, 48%).

IR ν (cm^{-1}) 3270, 3161, 3037, 2935, 2835, 1686. **^{1}H NMR** (500 MHz, CD_3OD) δ 8.59 (s, 1H, Ar*H*), 8.21 (s, 1H, Ar*H*), 6.28 (dd, *J* = 7.5, 5.9 Hz, 1H, 1'-*H*), 4.44 (dt, *J* = 6.5, 3.2 Hz, 1H, 3'-*H*), 4.32 (t, *J* = 7.0 Hz, 2H, C*H*$_2$), 4.05–4.01 (m, 7H, Cp*H*), 3.96 (q, *J* = 3.0 Hz, 1H, 4'-*H*), 3.90 (dd, *J* = 12.3, 3.0 Hz, 1H, 5'-*H*), 3.76–3.71 (m, 3H, Cp*H*, 5'-*H*), 2.39–2.30 (m, 3H, C*H*$_2$, 2'-*H*), 2.17 (ddd, *J* = 13.7, 7.5, 5.9 Hz, 1H, 2'-*H*), 2.08 (p, *J* = 7.0 Hz, 2H, C*H*$_2$). **^{13}C NMR** (126 MHz, CD_3OD) δ 170.8 (*C*O), 135.4 (Ar*C*), 121.6 (Ar*C*), 104.5 (Ar*C*), 87.8 (4'-*C*), 86.9 (*Cp*C), 85.8 (1'-*C*), 70.7 (Cp*C*H), 69.6 (Cp*C*H), 67.6 (3'-*C*), 67.1 (Cp*C*H), 66.5 (5'-*C*), 60.4 (*C*H_2), 40.1(2'-*C*), 30.4 (*C*H_2), 25.3 (*C*H_2). **HRMS** (ESI) calculated for $C_{24}H_{27}FeN_5O_5Na$ $[M+Na]^+$: *m/z* 544.1259, found: 544.1276.

General procedure D for the copper catalysed azide-alkyne cycloaddition of uridines and ferrocenes

To a stirred solution of 5-ethynyl-2'-deoxyuridine **27** (1 equiv.) and the corresponding azide (1 equiv.) in EtOAc (5 mL/mmol) and H_2O (1 mL/mmol) was added $CuSO_4 \cdot 5H_2O$ (10 mol%) and sodium ascorbate (20 mol%) and the yellow solution was allowed to stir at rt for 24 hrs. The solution was diluted with CH_2Cl_2 (20 mL/mmol) and H_2O (10 mL/mmol), and the aqueous layer was extracted with CH_2Cl_2 (3 x 20 mL/mmol). The combined organics were washed with $brine_{(sat.)}$ (10 mL/mmol), dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude solids were purified *via* silica gel chromatography (petrol/EtOAc/Et_3N 50:50:2→0:100:2) to yield the desired compounds.

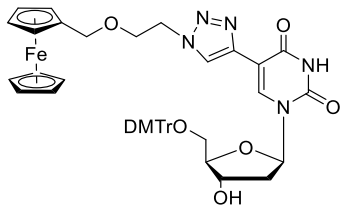
5-((3-ferrocenylpropyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl) phenylmethyl)-2'-deoxyuridine 63



General procedure C was followed using **27** (4.48 g, 8.3 mmol), **39** (2.69 g, 8.3 mmol), $CuSO_4 \cdot 5H_2O$ (207 mg, 10 mol%), sodium ascorbate (328 mg, 20 mol%) and EtOAc/H_2O (5:1) (50 mL) to yield the title compound as a yellow foam (5.47 g, 81%).

MP 103–107 °C. **IR** ν (cm^{-1}) 3159, 2981, 1684. **^{1}H NMR** (300 MHz, $(CD_3)_2SO$) δ 11.77 (s, 1H, N*H*), 8.37 (s, 1H, Ar*H*), 8.34 (s, 1H, Ar*H*), 7.37 (d, *J* = 7.2 Hz, 2H, Ar*H*), 7.30–7.20 (m, 6H, Ar*H*), 7.15 (t, *J* = 7.2 Hz, 1H, Ar*H*), 6.89–6.79 (m, 4H, Ar*H*), 6.19 (t, *J* = 6.5 Hz, 1H, 1'-*H*), 4.41 (t, *J* = 6.8 Hz, 2H, C*H*$_2$), 4.19 (dd, *J* = 9.6, 4.2 Hz, 1H, 3'-*H*), 4.09 (s, 2H, Cp*H*), 4.07 (s, 5H, Cp*H*), 4.05 (s, 2H, Cp*H*), 3.95 (q, *J* = 4.2 Hz, 1H, 4'-*H*), 3.70 (s, 3H, OC*H*$_3$), 3.70 (s, 3, OC*H*$_3$), 3.42–3.29 (m, 2H, 5'-*H*), 2.39–2.19 (m, 4H, 2'-*H*, C*H*$_2$), 1.74 (p, *J* = 6.9 Hz, 2H, C*H*$_2$). **^{13}C NMR** (126 MHz, CD_3CN) δ 162.1 (*C*O), 159.5 (Ar*C*), 159.5 (Ar*C*), 150.6 (Ar*C*), 146.0 (Ar*C*), 140.0 (Ar*C*), 136.9 (Ar*C*), 136.8 (Ar*C*), 136.6 (Ar*C*), 131.0 (Ar*C*), 131.0 (Ar*C*), 129.0 (Ar*C*), 128.8 (Ar*C*), 127.7 (Ar*C*), 123.0 (Ar*C*), 114.0 (Ar*C*), 114.0 (Ar*C*), 87.3 (4'-*C*), 87.1 (O*C*Ar$_3$), 86.6 (1'-*C*), 72.1 (3'-*C*), 69.5 (Cp*H*), 68.9 (Cp*H*), 68.2 (Cp*H*), 64.6 (5'-*C*), 55.9 (O*C*H$_3$), 50.6 (*C*H$_2$), 41.2 (2'-*C*), 32.1 (*C*H$_2$), 26.9 (*C*H$_2$). **HRMS** (ESI) calculated for $C_{45}H_{45}FeN_5NaO_7$ $[M+Na]^+$: *m/z* 846.2566, found 846.2593.

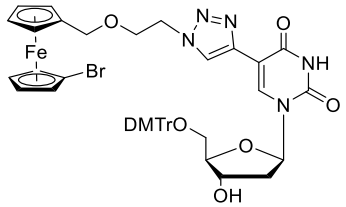
5-((2-(ferrocenylmethoxy)ethyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridine 64



General procedure C was followed using **27** (138 mg, 0.25 mmol), **56** (72 mg, 0.25 mmol), $CuSO_4 \cdot 5H_2O$ (6 mg, 10 mol%), sodium ascorbate (10 mg, 20 mol%), EtOAc (1 mL), and H_2O (0.2 mL) to yield the title compound as a yellow foam (92 mg, 44%).

MP 101–105 °C. **IR** ν (cm^{-1}) 3150, 2986, 2850, 1679. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.40 (s, 1H, Ar*H*), 8.23 (s, 1H, Ar*H*), 7.45–7.37 (m, 2H, Ar*H*), 7.35–7.28 (m, 4H, Ar*H*), 7.24–7.16 (m, 2H, Ar*H*), 7.15–7.08 (m, 1H, Ar*H*), 6.78 (d, *J* = 8.9 Hz, 4H, Ar*H*), 6.36–6.28 (m, 1H, 1'-*H*), 4.43 (t, *J* = 5.5 Hz, 2H, CH_2), 4.37 (dt, *J* = 7.4, 3.9 Hz, 1H, 3'-*H*), 4.25 (s, 2H, FcCH_2), 4.15 (t, *J* = 1.9 Hz, 2H, Cp*H*), 4.09 (t, *J* = 1.9 Hz, 2H, Cp*H*), 4.05 (s, 5H, Cp*H*), 4.05–4.01 (m, 1H, 4'-*H*), 3.76 (t, *J* = 5.5, 2H, CH_2), 3.75 (s, 6H, 2 x OCH_3), 3.42 (dd, *J* = 10.3, 4.9 Hz, 1H, 5'-*H*), 3.37 (dd, *J* = 10.3, 4.9, 1H, 5'-*H*), 2.49–2.38 (m, 1H, 2'-*H*), 2.30–2.21 (m, 1H, 2'-*H*). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 161.5 (*C*O), 158.5 (Ar*C*), 150.1 (Ar*C*), 144.7 (Ar*C*), 138.7 (Ar*C*), 135.7 (Ar*C*), 135.4 (Ar*C*), 130.1 (Ar*C*), 128.0 (Ar*C*), 127.9 (Ar*C*), 126.8 (Ar*C*), 123.0 (Ar*C*), 113.3 (Ar*C*), 86.8 (*C*Ar_3), 85.8 (4'-*C*), 85.5 (1'-*C*), 82.5 (Cp*C*), 71.7 (3'-*C*), 69.5 (Fc*C*H_2), 69.4 (Cp*C*H), 68.6 (Cp*C*H), 68.5 (Cp*C*H), 67.9 (*C*H_2), 63.7 (5'-*C*), 55.1 (O*C*H_3), 50.1 (*C*H_2), 41.5 (2'-*C*). **HRMS** (ESI) calculated for $C_{45}H_{45}FeN_5O_8Na$ $[M+Na]^+$: *m/z* 862.2510, found 862.2515.

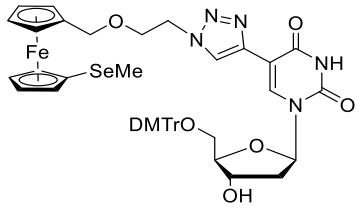
5-((2-((1'-bromoferrocenyl)methoxy)ethyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridine 65



General procedure C was followed using **27** (277 mg, 0.5 mmol), **39** (182 mg, 0.5 mmol), $CuSO_4 \cdot 5H_2O$ (12 mg, 10 mol%), sodium ascorbate (20 mg, 20 mol%), EtOAc (2 mL) and H_2O (0.5 mL) to yield the title compound as a yellow foam (221 mg, 48%).

MP 114–116 °C. **IR** ν (cm^{-1}) 3153, 2982, 1681. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.45 (s, 1H, Ar*H*), 8.28 (s, 1H, Ar*H*), 7.42 (d, *J* = 7.1 Hz, 2H, Ar*H*), 7.32 (d, *J* = 8.7 Hz, 4H), 7.22 (t, *J* = 7.6 Hz, 2H, Ar*H*), 7.13 (t, *J* = 7.3 Hz, 1H, Ar*H*), 6.80 (d, *J* = 8.9 Hz, 4H, Ar*H*), 6.33 (t, *J* = 6.7 Hz, 1H, 1'-*H*), 4.46 (t, *J* = 5.4 Hz, 2H, CH_2), 4.45–4.39 (m, 1H, 3'-*H*), 4.30 (t, *J* = 1.9 Hz, 2H, Cp*H*), 4.29 (s, 2H, FcCH_2), 4.18 (t, *J* = 1.9 Hz, 2H, Cp*H*), 4.17–4.15 (m, 2H, Cp*H*), 4.12–4.05 (m, 1H, 4'-*H*), 4.00 (t, *J* = 1.9 Hz, 2H, Cp*H*), 3.79 (t, *J* = 5.4 Hz, 2H, CH_2), 3.73 (s, 6H, 2 x OCH_3), 3.46 (dd, *J* = 10.3, 4.8 Hz, 1H, 5'-*H*), 3.39 (dd, *J* = 10.3, 4.8 Hz, 1H, 5'-*H*), 2.47 (ddd, *J* = 13.8, 6.7, 3.5 Hz, 1H, 2'-*H*), 2.28 (dt, *J* = 13.8, 6.7 Hz, 1H, 2'-*H*). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 161.5 (*C*O), 158.5 (Ar*C*), 158.4 (Ar*C*), 149.9 (Ar*C*), 144.7 (Ar*C*), 138.8 (Ar*C*), 136.1 (Ar*C*), 135.8 (Ar*C*), 135.8 (Ar*C*), 130.1 (Ar*C*), 130.1 (Ar*C*), 128.1 (Ar*C*), 127.9 (Ar*C*), 126.8 (Ar*C*), 123.3 (Ar*C*), 113.3 (Ar*C*), 113.2 (Ar*C*), 86.7 (*C*Ar_3), 86.1 (4'-*C*), 85.8 (1'-*C*), 84.1 (Cp*C*), 77.9 (Cp*C*), 72.2 (3'-*C*), 71.7 (Cp*C*H), 71.2 (Cp*C*H), 70.7 (Cp*C*H), 68.8 (Fc*C*H_2), 67.9 (*C*H_2), 67.7 (Cp*C*H), 63.9 (5'-*C*), 55.2 (O*C*H_3), 50.3 (*C*H_2), 40.2 (2'-*C*). **HRMS** (ESI) calculated for $C_{45}H_{44}BrFeN_5O_8Na$ $[M+Na]^+$: *m/z* 940.1615, found 941.1607.

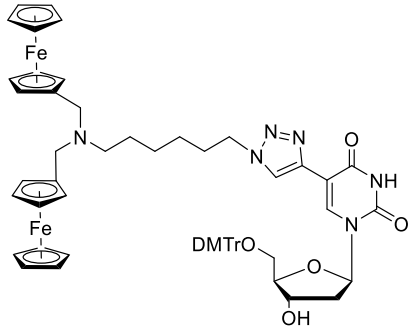
5-((2-((1'-(methylseleno)ferrocenyl)methoxy)ethyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridine 67



General procedure C was followed using **27** (277 mg, 0.5 mmol), **58** (189 mg, 0.5 mmol), $CuSO_4 \cdot 5H_2O$ (12 mg, 10 mol%), sodium ascorbate (20 mg, 20 mol%), EtOAc (2 mL) and H_2O (0.5 mL) to yield the title compound as a yellow foam (133 mg, 28%).

MP 108–109 °C. **IR** ν (cm^{-1}) 3160, 2983, 2868, 1684. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.43 (s, 1H, Ar*H*), 8.26 (s, 1H, Ar*H*), 7.43–7.39 (m, 2H, Ar*H*), 7.33–7.29 (m, 4H, Ar*H*), 7.21 (d, *J* = 7.7 Hz, 2H, Ar*H*), 7.16–7.11 (m, 1H, Ar*H*), 6.80 (d, *J* = 8.9 Hz, 4H, Ar*H*), 6.32 (t, *J* = 6.7 Hz, 1H, 1'-*H*), 4.46 (t, *J* = 5.4 Hz, 2H, C*H*$_2$), 4.42 (dt, *J* = 7.0, 3.8 Hz, 1H, 3'-*H*), 4.29 (s, 2H, C*H*$_2$), 4.22 (t, *J* = 1.8 Hz, 1H, Cp*H*), 4.16 (t, *J* = 1.8 Hz, 2H, Cp*H*), 4.13 (t, *J* = 1.8 Hz, 2H, Cp*H*), 4.10 (t, *J* = 1.8 Hz, 2H, Cp*H*), 4.09–4.05 (m, 1H, 4'-*H*), 3.78 (t, *J* = 5.4 Hz, 2H, C*H*$_2$), 3.73 (s, 6H, 2 x OC*H*$_3$), 3.46 (dd, *J* = 10.2, 4.8 Hz, 1H, 5'-*H*), 3.38 (dd, *J* = 10.2, 5.1 Hz, 1H, 5'-*H*), 2.47 (ddd, *J* = 13.9, 6.7, 3.8 Hz, 1H, 2'-*H*), 2.28 (dt, *J* = 13.9, 6.7 Hz, 1H, 2'-*H*), 2.08 (s, 3H, SeC*H*$_3$). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 161.3 (*C*O), 158.4 (Ar*C*), 158.4 (Ar*C*), 149.8 (Ar*C*), 144.8 (Ar*C*), 138.7 (Ar*C*), 136.5 (Ar*C*), 135.7 (Ar*C*), 135.7 (Ar*C*), 130.0 (Ar*C*), 130.0 (Ar*C*), 128.1 (Ar*C*), 127.9 (Ar*C*), 126.7 (Ar*C*), 123.0 (Ar*C*), 113.2 (Ar*C*), 113.2 (Ar*C*), 86.7 (*C*Ar$_3$), 85.7 (4'-*C*), 85.7 (1'-*C*), 83.3 (Cp*C*), 74.2 (Cp*C*), 73.3 (Cp*C*), 72.1 (3'-*C*), 70.6 (Cp*C*H), 70.0 (Cp*C*H), 69.7 (Cp*C*H), 69.1 (Fc*C*H$_2$), 67.8 (*C*H$_2$), 63.4 (5'-*C*), 55.1 (O*C*H$_3$), 50.2 (*C*H$_2$), 40.2 (2'-*C*), 9.8 (Se*C*H$_3$). **HRMS** (ESI) calculated for $C_{46}H_{47}FeN_5O_8SeNa$ $[M+Na]^+$: *m/z* 956.1831, found 956.1850.

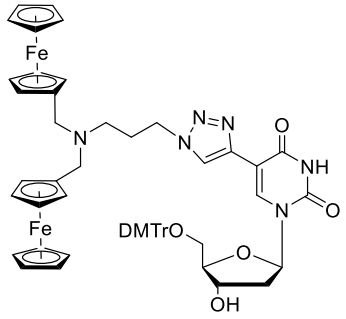
5-(1-(6-(bis(ferrocenylmethyl)amino)hexyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridine 85



General procedure C was followed using **27** (277 mg, 0.5 mmol), **78** (270 mg, 0.5 mmol), $CuSO_4 \cdot 5H_2O$ (12 mg, 10 mol%), sodium ascorbate (20 mg, 20 mol%) EtOAc (2 mL) and H_2O (0.5 mL) to yield the title compound as a yellow solid (212 g, 39%).

MP 93–94 °C. **IR** ν (cm^{-1}) 3078, 2930, 2858, 1687. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.45 (s, 1H, Ar*H*), 8.16 (s, 1H, Ar*H*), 7.42 (d, *J* = 7.0 Hz, 2H, Ar*H*), 7.32 (d, *J* = 7.7 Hz, 4H, Ar*H*), 7.23 (t, *J* = 7.6 Hz, 2H, Ar*H*), 7.14 (t, *J* = 7.2 Hz, 1H, Ar*H*), 6.80 (d, *J* = 8.9 Hz, 4H, Ar*H*), 6.33 (t, *J* = 6.1 Hz, 1H, 1'-*H*), 4.39 (dt, *J* = 7.0, 3.7 Hz, 1H, 3'-*H*), 4.26 (t, *J* = 7.2 Hz, 2H, C*H*$_2$), 4.16 (t, *J* = 1.8 Hz, 4H, Cp*H*), 4.10 (t, *J* = 1.8 Hz, 4H, Cp*H*), 4.08 (s, 10H, Cp*H*), 4.05 (q, *J* = 4.8 Hz, 1H, 4'-*H*), 3.47 (dd, *J* = 10.2, 4.8 Hz, 1H, 5'-*H*), 3.74 (s, 4H, FcC*H*$_2$), 3.38 (dd, *J* = 10.2, 4.8 Hz, 1H, 5'-*H*), 3.48–3.34 (m, 6H, 2 x OC*H*$_3$), 2.60 (q, *J* = 7.2 Hz, 2H, C*H*$_2$), 2.44 (ddd, *J* = 13.7, 6.1, 3.7 Hz, 1H, 2'-*H*), 2.30–2.22 (m, 3H, C*H*$_2$, 2'-*H*), 1.84 (p, *J* = 7.1 Hz, 2H, C*H*$_2$), 1.46–1.36 (m, 3H), 1.31–1.19 (m, 6H, C*H*$_2$), 1.05 (t, *J* = 7.2 Hz, 2H, C*H*$_2$). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 161.7 (*C*O), 158.5 (Ar*C*), 158.5 (Ar*C*), 149.9 (Ar*C*), 144.7 (Ar*C*), 138.9 (Ar*C*), 136.1 (Ar*C*), 135.8 (Ar*C*), 135.8 (Ar*C*), 130.1 (Ar*C*), 130.1 (Ar*C*), 128.2 (Ar*C*), 127.9 (Ar*C*), 126.8 (Ar*C*), 122.1 (Ar*C*), 113.3 (Ar*C*), 113.2 (Ar*C*), 86.7 (O*C*Ar$_3$), 86.0 (4'-*C*), 85.8 (1'-*C*), 83.4 (Cp*C*), 72.1 (3'-*C*), 70.3 (Cp*C*H), 68.6 (Cp*C*H), 67.9 (Cp*C*H), 63.9 (5'-*C*), 55.3 (O*C*H$_3$), 52.6 (Fc*C*H$_2$), 51.6 (*C*H$_2$), 50.3 (*C*H$_2$), 40.4 (2'-*C*), 30.3 (*C*H$_2$), 26.8 (*C*H$_2$), 26.8 (*C*H$_2$), 26.4 (*C*H$_2$). **HRMS** (ESI) calculated for $C_{60}H_{64}Fe_2N_6O_7Na$ $[M+Na]^+$: *m/z* 1115.3428, found 1115.3431.

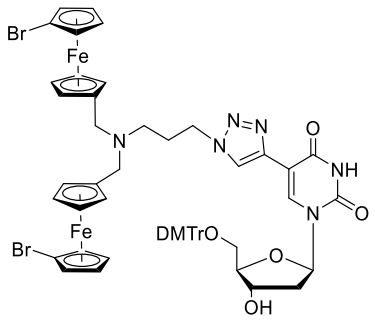
5-(1-(3-(bis(ferrocenylmethyl)amino)propyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridine 89



General procedure C was followed using **27** (277 mg, 0.5 mmol), **88** (247 mg, 0.5 mmol), $CuSO_4 \cdot 5H_2O$ (12 mg, 10 mol%), sodium ascorbate (20 mg, 20 mol%) in EtOAc (2.5 mL) and H_2O (2.5 mL) to yield the title compound as a yellow solid (611 mg, 65%).

MP 95–97 °C. **IR** ν (cm^{-1}) 3083, 2936, 2832, 1686, 1681. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.42 (s, 1H, Ar*H*), 8.08 (s, 1H, Ar*H*), 7.42–7.39 (m, 2H, Ar*H*), 7.32 (dd, *J* = 9.0, 1.3 Hz, 4H, Ar*H*), 7.24–7.20 (m, 2H, Ar*H*), 7.16–7.12 (m, 1H, Ar*H*), 6.82–6.77 (m, 4H, Ar*H*), 6.32 (t, *J* = 6.7 Hz, 1H, 1'-*H*), 4.41 (dt, *J* = 6.7, 3.5 Hz, 1H, 3'-*H*), 4.22 (t, *J* = 6.6, Hz, 2H, CH_2), 4.17–4.11 (m, 4H, Cp*H*), 4.09 (t, *J* = 1.9 Hz, 4H, Cp*H*), 4.07 (s, 10H, Cp*H*), 4.07–4.04 (m, 1H, 4'-*H*), 3.73 (s, 3H, OCH_3), 3.73 (s, 3H, OCH_3), 3.45 (dd, *J* = 10.3, 5.0 Hz, 1H, 5'-*H*), 3.43 (s, 4H, FcCH_2), 3.38 (dd, *J* = 10.3, 5.0 Hz, 1H, 5'-*H*), 2.45 (ddd, *J* = 13.7, 6.7, 3.5 Hz, 1H, 2'-*H*), 2.32 (t, *J* = 6.6 Hz, 2H, CH_2), 2.27 (dt, *J* = 13.7, 6.7 Hz, 1H, 2'-*H*), 1.96 (p, *J* = 6.6 Hz, 2H, CH_2). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 161.4 (*C*O), 158.5 (Ar*C*), 158.4 (Ar*C*), 149.8 (Ar*C*), 144.6 (Ar*C*), 138.6 (Ar*C*), 135.9 (Ar*C*), 135.8 (Ar*C*), 135.7 (Ar*C*), 130.1 (Ar*C*), 130.0 (Ar*C*), 128.1 (Ar*C*), 127.8 (Ar*C*), 126.8 (Ar*C*), 122.6 (Ar*C*), 113.2 (Ar*C*), 113.2 (Ar*C*), 86.7 (O*C*Ar_3), 85.9 (4'-*C*), 85.7 (1'-*C*), 83.2 (Cp*C*), 72.2 (3'-*C*), 70.1 (Cp*C*H), 68.5 (Cp*C*H), 68.0 (Cp*C*H), 63.8 (5'-*C*), 55.2 (O*C*H_3), 52.9 (Fc*C*H_2), 48.3 (*C*H_2), 48.3 (*C*H_2), 40.3 (2'-*C*), 28.1 (*C*H_2). **HRMS** (ESI) calculated for $C_{57}H_{58}Fe_2N_6NaO_7$ $[M+Na]^+$: *m/z* 1073.2958, found 1073.2961.

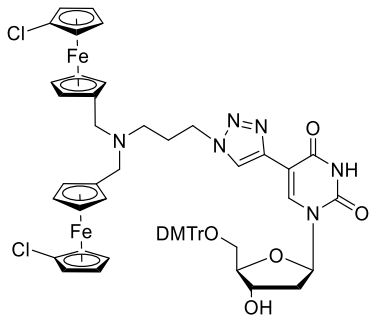
5-(1-(3-(bis((1'-bromo)ferrocenylmethyl)amino)propyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridine 103



General procedure C was followed using **27** (2.22 g, 4.0 mmol), **98** (2.50 g, 4.0 mmol), $CuSO_4 \cdot 5H_2O$ (100 mg, 10 mol%), sodium ascorbate (247 mg, 20 mol%) in EtOAc (20 mL) and H_2O (4 mL) to yield the title compound as a yellow solid (3.05 g, 74%).

MP 98–101 °C. **IR** ν (cm^{-1}) 2948, 2828, 1677. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.43 (s, 1H, Ar*H*), 8.09 (s, 1H, Ar*H*), 7.46–7.37 (m, 2H, Ar*H*), 7.32 (d, *J* = 8.1 Hz, 4H, Ar*H*), 7.23 (t, *J* = 7.7 Hz, 2H, Ar*H*), 7.17–7.11 (m, 1H, Ar*H*), 6.80 (d, *J* = 8.9 Hz, 4H, Ar*H*), 6.33 (t, *J* = 6.7 Hz, 1H, 1'-*H*), 4.42 (dt, *J* = 7.0, 3.7 Hz, 1H, 3'-*H*), 4.30 (t, *J* = 1.9 Hz, 4H, Cp*H*), 4.28–4.22 (m, 2H, CH_2), 4.19–4.13 (m, 8H, Cp*H*), 4.09–4.06 (m, 1H, 4'-*H*), 4.02 (t, *J* = 1.9 Hz, 4H, Cp*H*), 3.74 (s *3*H, OCH_3), 3.73 (s *3*H, OCH_3), 3.48–3.43 (m, 5H, FcCH_2, 5'-*H*), 3.39 (dd, *J* = 10.3, 5.1 Hz, 1H, 5'-*H*), 2.51–2.42 (m, 1H, 2'-*H*), 2.35 (t, *J* = 6.8 Hz, 2H, CH_2), 2.28 (dt, *J* = 13.7, 6.7 Hz, 1H, 2'-*H*), 1.98 (p, *J* = 6.8 Hz, 2H, CH_2). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 161.4 (*C*O), 158.5 (Ar*C*), 149.9 (Ar*C*), 144.7 (Ar*C*), 138.6 (Ar*C*), 136.0 (*C*O), 135.8 (Ar*C*), 135.7 (Ar*C*), 130.1 (Ar*C*), 128.1 (Ar*C*), 127.9 (Ar*C*), 126.8 (Ar*C*), 122.7 (Ar*C*), 113.3 (Ar*C*), 86.7 (O*C*Ar_3), 85.9 (4'-*C*), 85.7 (1'-*C*), 84.8 (Cp*C*), 78.1 (Cp*C*), 72.6 (Cp*C*H), 72.3 (3'-C), 70.7 (Cp*C*H), 70.6 (Cp*C*H), 67.7 (Cp*C*H), 63.8 (5'-*C*), 55.2 (O*C*H_3), 52.1 (Fc*C*H_2), 48.6 (*C*H_2), 48.3 (*C*H_2), 40.3 (2'-*C*), 28.1 (*C*H_2). **HRMS** (ESI) calculated for $C_{57}H_{57}Br_2Fe_2N_6O_7$ $[M+H]^+$: *m/z* 1209.1341, found 1209.1428.

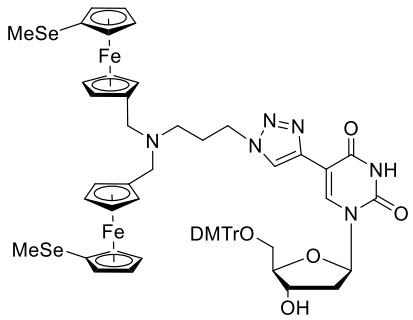
5-(1-(3-(bis(chloroferrocenylmethyl)amino)propyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridine 104



General procedure C was followed using compound **27** (831 mg, 1.5 mmol), **101** (834 mg, 1.5 mmol), $CuSO_4 \cdot 5H_2O$ (38 mg, 10 mol%), sodium ascorbate (60 mg, 20 mol%) in EtOAc (7.5 mL) and H_2O (1.5 mL) to yield the title compound as a yellow solid (896 mg, 53%).

MP 97–100 °C. **IR** ν (cm^{-1}) 2956, 2834, 1690. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.40 (s, 1H, Ar*H*), 8.04 (s, 1H, Ar*H*), 7.44–7.38 (m, 2H, Ar*H*), 7.32 (d, *J* = 8.6 Hz, 4H, Ar*H*), 7.26–7.22 (m, 2H, Ar*H*), 7.18–7.12 (m, 1H, Ar*H*), 6.81 (d, *J* = 8.6 Hz, 4H, Ar*H*), 6.31 (t, *J* = 6.6 Hz, 1H, 1'-*H*), 4.41 (dt, *J* = 7.4, 4.1 Hz, 1H, 3'-*H*), 4.28 (t, *J* = 1.9 Hz, 4H, Cp*H*), 4.25 (t, *J* = 7.2 Hz, 2H, CH_2) 4.20–4.13 (m, 6H, Cp*H*), 4.08–4.00 (m, 1H, 4'-*H*), 3.98 (t, *J* = 1.9 Hz, 2H, Cp*H*), 3.76 (s, 6H, 2 x CH_3), 3.51–3.46 (m, 1H, 5'-*H*), 3.43 (s, 4H, FcCH_2), 3.42–3.36 (m, 1H, 5'-*H*), 2.45 (ddd, *J* = 13.8, 6.6, 4.1 Hz, 1H, 2'-*H*), 2.37–2.26 (m, 3H, CH_2, 2'-*H*), 1.97 (p, *J* = 6.8 Hz, 2H, CH_2). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 161.2 (*C*O), 158.6 (Ar*C*), 149.7 (Ar*C*), 144.7 (Ar*C*), 138.6 (Ar*C*), 135.9 (*C*O), 135.8 (Ar*C*), 130.2 (Ar*C*), 128.2 (Ar*C*), 128.0 (Ar*C*), 126.9 (Ar*C*), 122.7 (Ar*C*), 113.3 (Ar*C*), 86.9 (O*C*Ar$_3$), 85.7 (4'-*C*), 85.7 (1'-*C*), 84.9 (Cp*C*), 77.9 (Cp*C*), 72.5 (3'-*C*), 72.3 (Cp*C*H), 70.3 (Cp*C*H), 68.5 (Cp*C*H), 66.7 (Cp*C*H), 63.9 (5'-*C*), 55.3 (OCH_3), 52.3 (Fc*C*H_2), 48.7 (*C*H_2), 48.4 (CH_2), 40.3 (2'-*C*), 28.3 (*C*H_2). **HRMS** (ESI) calculated for $C_{57}H_{57}Cl_2Fe_2N_6O_7$ $[M+H]^+$: *m/z* 1119.2363, found 1119.2405

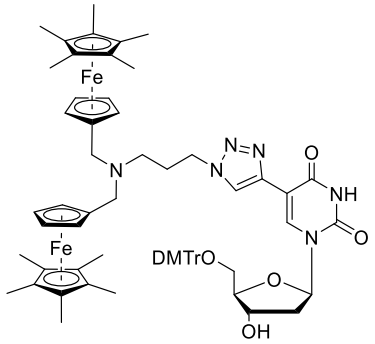
5-(1-(3-(bis((1'-methylseleno)ferrocenylmethyl)amino)propyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridine 105



General procedure C was followed using **27** (249 mg, 0.47 mmol), **99** (325 mg, 0.47 mmol), $CuSO_4 \cdot 5H_2O$ (12 mg, 10 mol%), sodium ascorbate (20 mg, 20 mol%) EtOAc (1.5 mL) and H_2O (0.5 mL) to yield the title compound as a yellow foam (308 mg, 50%).

MP 93–95 °C. **IR** ν (cm^{-1}) 2949, 2832, 1697. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.41 (s, 1H, Ar*H*), 8.07 (s, 1H, Ar*H*), 7.41 (d, *J* = 8.2 Hz, 2H, Ar*H*), 7.32 (d, *J* = 8.9 Hz, 4H, Ar*H*), 7.22 (t, *J* = 7.7 Hz, 2H, Ar*H*), 7.16–7.11 (m, 1H, Ar*H*), 6.80 (d, *J* = 8.9 Hz, 4H, Ar*H*), 6.32 (t, *J* = 6.6 Hz, 1H, 1'-*H*), 4.41 (dt, *J* = 7.1, 3.5 Hz, 1H, 3'-*H*), 4.25–4.20 (m, 6H, C*H*$_2$, Cp*H*), 4.16–4.10 (m, 8H, Cp*H*), 4.10 (t, *J* = 1.8 Hz, 4H, Cp*H*), 4.09–4.02 (m, 1H, 4'-*H*), 3.74 (s, 6H, 2 x OC*H*$_3$), 3.46 (dd, *J* = 10.2, 5.1 Hz, 1H, 5'-*H*), 3.43 (s, 4H, FcC*H*$_2$), 3.38 (dd, *J* = 10.2, 5.1 Hz, 1H, 5'-*H*), 2.46 (ddd, *J* = 13.7, 6.6, 3.5 Hz, 1H, 2'-*H*), 2.33 (t, *J* = 6.6 Hz, 2H, C*H*$_2$), 2.29–2.24 (m, 1H, 2'-*H*), 2.09 (s, 6H, SeC*H*$_3$), 1.96 (p, *J* = 6.6 Hz, 2H, C*H*$_2$). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 161.3 (*C*O), 158.3 (Ar*C*), 149.6 (Ar*C*), 144.8 (Ar*C*), 138.3 (Ar*C*), 135.9 (*C*O), 135.7 (Ar*C*), 130.1 (Ar*C*), 128.1 (Ar*C*), 127.8 (Ar*C*), 126.7 (Ar*C*), 122.5 (Ar*C*), 113.2 (Ar*C*), 86.7 (O*C*Ar$_3$), 86.6 (4'-*C*), 85.7 (1'-*C*), 84.1 (Cp*C*H), 74.2 (Cp*C*), 73.1 (Cp*C*H), 71.5 (Cp*C*H), 71.5 (3'-C), 69.9 (Cp*C*H), 69.4 (Cp*C*), 63.8 (5'-*C*), 55.2 (O*C*H$_3$), 52.3 (Fc*C*H$_2$), 48.3 (*C*H$_2$), 48.2 (*C*H$_2$), 40.2 (2'-*C*), 28.1 (*C*H$_2$), 9.9 (Se*C*H$_3$). **HRMS** (ESI) calculated for $C_{59}H_{63}Fe_2N_6O_7Se_2$ $[M+H]^+$: *m/z* 1237.1813, found 1237.1803.

5-(1-(3-(bis((1',2',3',4',5'-pentamethyl)ferrocenylmethyl)amino)propyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridine 107



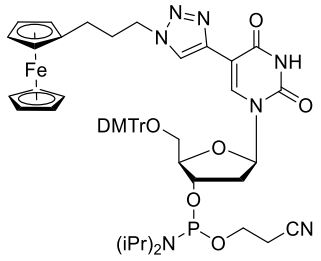
General procedure C was followed using **27** (720 mg, 1.3 mmol), **102** (826 mg, 1.3 mmol) $CuSO_4 \cdot 5H_2O$ (32 mg, 0.13 mmol), sodium ascorbate (26 mg, 0.13 mmol) and EtOAc/H_2O (4:1) (5 mL) to yield the title compound as a yellow solid (1.28 g, 82%).

MP 85–88 °C. **IR** ν (cm^{-1}) 2985, 2856, 1687. **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.37 (s, 1H, Ar*H*), 8.01 (s, 1H, Ar*H*), 7.44–7.37 (m, 2H, Ar*H*), 7.32 (d, *J* = 8.3 Hz, 4H, Ar*H*), 7.23 (t, *J* = 7.6 Hz, 2H, Ar*H*), 7.15 (t, *J* = 7.3 Hz, 1H, Ar*H*), 6.81 (d, *J* = 8.8 Hz, 4H, Ar*H*), 6.32 (t, *J* = 6.6 Hz, 1H, 1'-*H*), 4.39 (dt, *J* = 7.4, 4.0 Hz, 1H, 3'-*H*), 4.19 (t, *J* = 7.0 Hz, 2H, CH_2), 4.04–3.98 (m, 1H, 4'-*H*), 3.75 (s, 6H, OCH_3), 3.64 (t, *J* = 1.8 Hz, 4H, Cp*H*), 3.55 (s, 4H, Cp*H*), 3.48 (dd, *J* = 10.2, 4.7 Hz, 1H, 5'-*H*), 3.37 (dd, *J* = 10.2, 5.5 Hz, 1H, 5'-*H*), 3.24 (s, 4H, FcCH_2), 2.44 (ddd, *J* = 13.7, 6.6, 4.0 Hz, 1H, 2'-*H*), 2.35–2.21 (m, 3H, 2'-*H*, CH_2), 1.90 (p, *J* = 7.0 Hz, 2H, CH_2), 1.85 (s, 30H, 10 x FcCH_3). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 161.1 (*C*O), 158.6 (Ar*C*), 158.6 (Ar*C*), 149.7 (Ar*C*), 144.7 (Ar*C*), 138.6 (Ar*C*), 135.8 (Ar*C*), 135.8 (Ar*C*), 130.2 (Ar*C*), 130.2 (Ar*C*), 128.2 (Ar*C*), 128.0 (Ar*C*), 126.9 (Ar*C*), 122.7 (Ar*C*), 113.4 (Ar*C*), 113.4 (Ar*C*), 86.9 (O*C*Ar$_3$), 85.6 (4'-*C*), 85.5 (1'-*C*), 82.0 (Cp*C*), 80.2 (Cp*C*), 72.6 (Cp*C*H), 72.5 (3'-*C*), 72.4 (Cp*C*H), 63.9 (5'-*C*), 55.3 (O*C*H$_3$), 55.3 (O*C*H$_3$), 52.2 (Fc*C*H$_2$), 48.6 (*C*H$_2$), 48.5 (*C*H$_2$), 40.3 (2'-*C*), 28.5 (*C*H$_2$), 11.4 (10 x Fc*C*H$_3$). **HRMS** (ESI) calculated for $C_{67}H_{79}Fe_2N_6O_7$ $[M+H]^+$: *m/z* 1191.4709, found 1191.4755.

General procedure D for phosphoramidite formation

To a stirred solution of alcohol (2.8 mmol) in anhydrous, deoxygenated THF (25 mL) under nitrogen was added DIPEA (1.95 mL, 11.2 mmol) and the solution was stirred for 10 minutes. 2-cyano-*N,N*-diisopropylchlorophosphoramidite (1 g, 4.2 mmol) was added dropwise *via* syringe and allowed to stir at rt. After 5 minutes, a precipitate formed, and the reaction was treated with H_2O (0.10 mL) and allowed to stir at rt for 30 minutes. The solution was treated with H_2O (0.20 mL) to give a homogenous solution, which was then treated with EtOAc/TEA (25 mL, 1:1) and poured into $NaHCO_{3(sat.)(aq.)}$ (10 mL) and the aqueous layer was extracted with EtOAc (3 x 10 mL). The combined organics were dried over $MgSO_4$, concentrated and purified *via* flash chromatography (petrol/EtOAc/TEA 40:60:5) under a stream of nitrogen to yield yellow solids/oils.

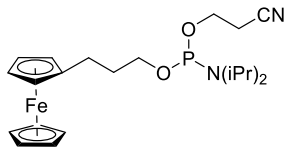
2-cyanoethyl *N,N*-diisopropyl-(3'-*O*-(5-(3-ferrocenylpropyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridinyl))-phosphoramidite 68



General procedure D was followed using **63** (2.26 g, 2.8 mmol) to yield the title compound as a mixture of diastereomers, as a yellow solid (2.45 g, 86%).

MP 103–105 °C. **IR** ν (cm^{-1}) 2966, 1684, 1507. **^{1}H NMR** (500 MHz, CD_3CN) δ 8.45 (s, 1H, Ar*H*), 8.17 (s, 1H, Ar*H*), 7.47–7.40 (m, 2H, Ar*H*), 7.36–7.30 (m, 4H, Ar*H*), 7.26–7.21 (m, 2H, Ar*H*), 7.19–7.14 (m, 1H, Ar*H*), 6.85–6.80 (m, 4H, Ar*H*), 6.26–6.18 (m, 1H, 1'-*H*), 4.53–4.45 (m, 1H, 3'-*H*), 4.34 (td, *J* = 7.0, 2.5 Hz, 2H, C*H*$_2$), 4.17 (q, *J* = 3.5 Hz, 1H, 4'-*H*), 4.07–4.06 (m, 2H, Cp*H*), 4.04 (s, 5H, Cp*H*), 4.03 (t, *J* = 1.9 Hz, 2H, Cp*H*), 3.83–3.75 (m, 1H, C*H*$_2$), 3.74–3.72 (m, 6H, 2 x OC*H*$_3$), 3.69–3.61 (m, 1H, C*H*$_2$), 3.61–3.52 (m, 2H, NC*H*(CH$_3$)$_2$), 3.41–3.35 (m, 1H, 5'-*H*), 3.35–3.28 (m, 1H, 5'-*H*), 2.67–2.60 (m, 1H, C*H*$_2$), 2.56–2.45 (m, 2H, C*H*$_2$, 2'-*H*), 2.41–2.34 (m, 1H, 2'-*H*), 2.35–2.27 (m, 2H, C*H*$_2$), 2.12–2.05 (m, 2H, C*H*$_2$), 1.19–1.02 (m, 12H, 4 x C*H*$_3$). **^{13}C NMR** (126 MHz, CD_3CN) δ 162.1 (CO), 159.6 (Ar*C*), 150.6 (Ar*C*), 145.9 (Ar*C*), 139.9 (Ar*C*), 136.7 (Ar*C*), 136.7 (Ar*C*), 136.5 (Ar*C*), 131.0 (Ar*C*), 129.0 (Ar*C*), 128.8 (Ar*C*), 127.7 (Ar*C*), 123.0 (Ar*C*), 119.3 (*C*N), 114.1 (Ar*C*), 88.6 (O*C*Ar$_3$), 86.7 (5'-*C*), 86.4 (d, *J* = 4.3 Hz, 4'-*C*), 74.5 (d, *J* = 17.6 Hz, 3'-*C*), 69.3 (Cp*C*H), 68.9 (Cp*C*H), 68.1 (Cp*C*H), 64.3 (5'-*C*), 59.5 (d, *J* = 19.1 Hz, C*H*$_2$), 55.8 (O*C*H$_3$), 55.8 (O*C*H$_3$), 50.5 (C*H*$_2$), 44.0 (d, *J* = 12.4 Hz, N*C*H(CH$_3$)$_2$), 40.3 (d, *J* = 4.5 Hz, 2'-*C*), 32.0 (*C*H$_2$), 26.8 (*C*H$_2$), 24.9–24.7 (m, 4 x *C*H$_3$), 20.9 (d, *J* = 7.1 Hz, *C*H$_2$). **^{31}P NMR** (202 MHz, CD_3CN) δ 148.09. **HRMS** (ESI) calculated for $C_{54}H_{62}FeN_7NaO_8P$ $[M+Na]^+$: *m/z* 1046.3639, found 1046.3688.

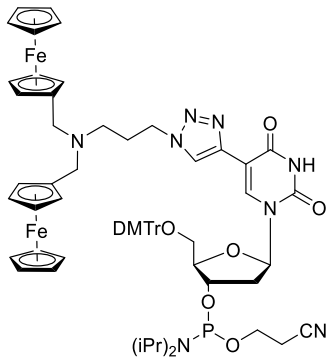
2-cyanoethyl (3-(ferrocenyl)propyl) *N,N*-diisopropylphosphoramidite 69[110]



General procedure D was followed using **38** (0.682 g, 2.8 mmol) to yield the title compound as a mixture of diastereomers, as a yellow oil (1.168 g, 93%).

^{1}H NMR (500 MHz, C_6D_6) δ 4.03 (s, 5H, Cp*H*), 4.01 (s, 2H, Cp*H*), 3.96 (s, 2H, Cp*H*), 3.74– 3.67 (m, 1H, POCH_2), 3.66–3.60 (m, 1H, POCH_2), 3.60–3.52 (m, 2H, NC*H*), 3.43–3.36 (m, 1H, POCH_2), 3.35–3.26 (m, 1H, POCH_2), 2.39 (t, *J* = 7.7 Hz, 2H, CH_2), 1.84–1.71 (m, 4H, CH_2, CH_2CN), 1.15 (t, *J* = 6.7 Hz, 12H, 4 x CH_3). **^{13}C NMR** (126 MHz, C_6D_6) δ 117.6 (*C*N), 88.7 (Cp*C*), 68.9 (Cp*C*H), 68.4 (d, *J* = 1.8 Hz, Cp*C*H), 67.6 (Cp*C*H), 63.5 (d, *J* = 16.3 Hz, PO*C*H_2), 58.6 (d, *J* = 18.7 Hz, PO*C*H_2), 43.4 (d, *J* = 12.4 Hz, PN*C*H), 33.1 (d, *J* = 7.0 Hz, *C*H_2), 26.3 (*C*H_2), 24.7 (d, *J* = 7.3 Hz, *C*H_3), 20.2 (d, *J* = 6.7 Hz, *C*H_2). **^{31}P NMR** (202 MHz, C_6D_6) δ 147.68.

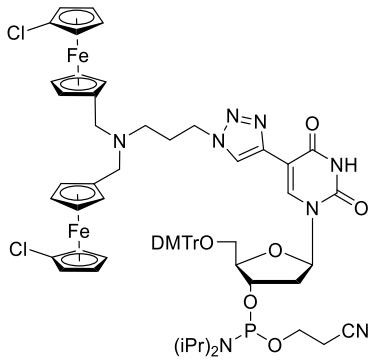
2-cyanoethyl *N,N*-diisopropyl-(3'-*O*-(5-(1-(3-(ferrocenylmethyl)amino)propyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridinyl))-phosphoramidite 108



General procedure D was followed using **89** (1.40 g, 1.3 mmol) to yield the title compound as a mixture of diastereomers, as a yellow foam (1.295 g, 80%).

MP 120–121 °C. **IR** ν (cm^{-1}) 3083, 2965, 2837, 2256, 1698, 1507. **^{1}H NMR** (500 MHz, CD_3CN) δ 9.36 (s, 1H, N*H*), 8.42 (s, 1H, Ar*H*), 8.06 (s, 1H, Ar*H*), 7.48–7.43 (m, 2H, Ar*H*), 7.39–7.32 (m, 4H, Ar*H*), 7.30–7.23 (m, 2H, Ar*H*), 7.22–7.17 (m, 1H, Ar*H*), 6.88–6.83 (m, 4H, Ar*H*), 6.25 (q, *J* = 6.7 Hz, 1H, 1'-*H*), 4.51 (dddt, *J* = 13.4, 9.9, 6.8, 3.4 Hz, 1H, 3'-*H*), 4.28 (td, *J* = 7.0, 2.5 Hz, 2H, CH_2), 4.20–4.17 (m, 1H, 4'-*H*), 4.16 (s, 2H, Cp*H*), 4.09 (s, 10H, Cp*H*), 4.08 (s, 4H, Cp*H*), 3.85 – 3.78 (m, 1H, CNCH_2), 3.76 (s, 6H, 2 x OCH_3), 3.67 (dtd, *J* = 8.0, 6.0, 1.9 Hz, 1H, CNCH_2), 3.64–3.54 (m, 2H, PNC*H*), 3.42–3.37 (m, 5H, FcCH_2, 5'-*H*), 3.38–3.28 (m, 1H, 5'-*H*), 2.68–2.52 (m, 2H, POCH_2), 2.49 (ddd, *J* = 13.7, 6.7, 3.4 Hz, 1H, 2'-*H*), 2.43–2.33 (m, 1H, 2'-*H*), 2.28 (t, *J* = 6.7 Hz, 2H, CH_2), 1.21–1.05 (m, 12H, 4 x CH_3). **^{13}C NMR** (126 MHz, CD_3CN) δ 161.9 (*C*O), 159.6 (Ar*C*), 150.5 (Ar*C*), 145.9 (Ar*C*), 139.7 (Ar*C*), 136.7 (Ar*C*), 136.3 (Ar*C*), 131.0 (Ar*C*), 129.0 (Ar*C*), 128.8 (Ar*C*), 127.7 (Ar*C*), 123.4 (Ar*C*), 119.4 (*C*N), 114.1 (Ar*C*), 87.4 (O*C*Ar$_3$), 86.6 (1'-*C*), 86.3 (d, *J* = 4.4 Hz, 4'-*C*), 84.9 (Cp*C*), 74.0 (d, *J* = 17.1 Hz, 3'-*C*), 70.9 (Cp*C*H), 69.4 (Cp*C*H), 68.6 (Cp*C*H), 64.3 (5'-*C*), 59.6 (d, *J* = 2.8 Hz, PO*C*H$_2$), 55.9 (O*C*H$_3$), 55.9 (O*C*H$_3$), 53.4 (Fc*C*H$_2$), 49.6 (*C*H$_2$), 48.8 (*C*H$_2$), 44.0 (d, *J* = 12.4 Hz, N*C*H(CH$_3$)$_2$), 40.3 (d, *J* = 3.7 Hz, 2'-*C*), 28.7 (*C*H$_2$), 25.2–24.6 (m, NCH(*C*H$_3$)$_2$), 21.0 (d, *J* = 7.3 Hz, *C*H$_2$CN). **^{31}P NMR** (202 MHz, CD_3CN) δ 148.11. **HRMS** (ESI) calculated for $C_{66}H_{75}Fe_2N_8O_8PNa$ $[M+Na]^+$: *m/z* 1273.4041, found 1273.4099.

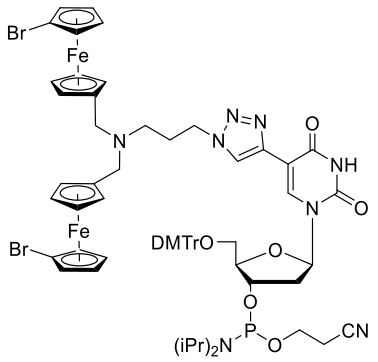
2-cyanoethyl *N,N*-diisopropyl-(3'-*O*-(5-(1-(3-(bis(chloroferrocenylmethyl)amino) propyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridinyl))-phosphoramidite 109



General procedure D was followed using **104** (896 mg, 0.8 mmol) to yield the title compound as a mixture of diastereomers, as a yellow powder (480 mg, 45%).

MP 125–128 °C. **IR** ν (cm^{-1}) 2963, 2233, 1684. **^{1}H NMR** (500 MHz, CD_3CN) δ 8.37 (s, 1H, Ar*H*), 8.04 (s, 1H, Ar*H*), 7.46–7.39 (m, 2H, Ar*H*), 7.36–7.28 (m, 4H, Ar*H*), 7.27–7.20 (m, 2H, Ar*H*), 7.20–7.12 (m, 1H, Ar*H*), 6.85–6.78 (m, 4H, Ar*H*), 6.25–6.17 (m, 1H, 1'-*H*), 4.54–4.42 (m, 1H, 3'-*H*), 4.37 (m, 2H, CH_2) 4.30 (t, *J* = 1.9 Hz, 4H, Cp*H*), 4.27 (td, *J* = 6.9, 3.6 Hz, 1H, 4'-*H*), 4.18 (t, *J* = 1.9 Hz, 4H, Cp*H*), 4.14 (t, *J* = 1.9 Hz, 4H, Cp*H*), 4.02 (t, *J* = 1.9 Hz, 4H, Cp*H*), 3.83–3.75 (m, 1H, *C*H_2), 3.73 (s, 6H, 3 x O*C*H_3), 3.71–3.60 (m, 1H, *C*H_2), 3.63–3.49 (m, 2H, NC*H*$(CH_3)_2$), 3.38 (s, 4H, Fc*C*H_2), 3.33 (dd, *J* = 10.6, 3.6 Hz, 1H, 5'-*H*), 3.29 (dd, *J* = 10.6, 5.2 Hz, 1H, 5'-*H*), 2.63 (td, *J* = 6.0, 1.9 Hz, 1H, C*H*$_2$), 2.51 (p, *J* = 6.0 Hz, 1H, C*H*$_2$), 2.45 (ddd, *J* = 13.8, 6.1, 3.5 Hz, 1H, 2'-*H*), 2.34 (dtd, *J* = 13.8, 6.1, 3.5 Hz, 1H, 2'-*H*), 2.27 (t, *J* = 6.0 Hz, 2H, C*H*$_2$), 2.02–1.99 (m, 2H, C*H*$_2$), 1.16–1.02 (m, 12H, 4 x C*H*$_3$). **^{13}C NMR** (126 MHz, CD_3CN) δ 161.0 (*C*O), 158.6 (Ar*C*), 149.6 (Ar*C*), 145.0 (Ar*C*), 138.7 (Ar*C*), 135.8 (Ar*C*), 135.7 (Ar*C*), 130.1 (Ar*C*), 130.0 (Ar*C*), 128.0 (Ar*C*), 127.8 (Ar*C*), 122.5 (Ar*C*), 118.4 (*C*N), 113.1 (Ar*C*), 86.4 (O*C*Ar$_3$), 85.5 (4'-*C*), 85.3 (1'-*C*), 85.1 (Cp*C*), 73.0 (d, *J* = 17.1 Hz, 3'-*C*), 72.2 (Cp*C*H), 70.8 (Cp*C*), 70.0 (Cp*C*H), 68.2 (Cp*C*H), 66.8 (Cp*C*H), 63.4 (5'-*C*), 58.53 (d, *J* = 19.2 Hz, PO*C*H_2), 54.9 (O*C*H_3), 54.9 (O*C*H_3), 51.6 (Fc*C*H_2), 48.7 (*C*H_2), 47.8 (*C*H_2), 43.05 (d, *J* = 12.3 Hz, N*C*H$(CH_3)_2$), 39.3 (2'-C), 27.7 (*C*H_2), 24.04–23.74 (m, NCH(*C*$H_3)_2$) 20.35–19.87 (m, *C*H_2CN). **^{31}P NMR** (202 MHz, CD_3CD) δ 148.11. **HRMS** (ESI) $C_{66}H_{73}Cl_2Fe_2N_8O_8PNa$ $[M+Na]^+$: *m/z* 1341.3250, found 1341.3250.

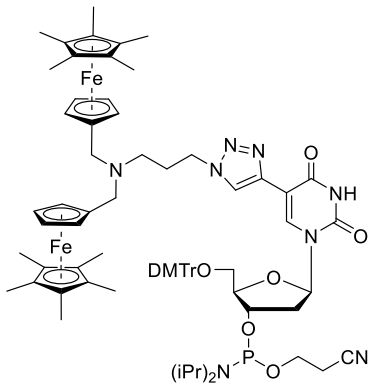
2-cyanoethyl *N,N*-diisopropyl-(3'-*O*-(5-(1-(3-(bis((1'-bromo)ferrocenylmethyl) amino)propyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridinyl))-phosphoramidite 110



General procedure D was followed using **103** (2.41 g, 2.0 mmol) to yield the title compound as a mixture of diastereomers, as a yellow powder (1.63 g, 58%).

MP 119–121 °C. **IR** ν (cm^{-1}) 3078, 2954, 2839, 2253, 1701. **^{1}H NMR** (500 MHz, CD_3CN) δ 8.44 (s, 1H, Ar*H*), 8.08 (s, 1H, Ar*H*), 7.50–7.43 (m, 2H, Ar*H*), 7.39–7.32 (m, 4H, Ar*H*), 7.29–7.24 (m, 2H, Ar*H*), 7.22–7.15 (m, 1H, Ar*H*), 6.88–6.80 (m, 4H, Ar*H*), 6.29–6.20 (m, 1H, 1'-*H*), 4.57–4.45 (m, 1H, 3'-*H*), 4.33 (t, *J* = 1.8 Hz, 4H, Cp*H*), 4.34–4.27 (m, 2H, CH_2), 4.25–4.20 (m, 1H, 4'-*H*), 4.18 (t, *J* = 1.8 Hz, 4H, Cp*H*), 4.16 (t, *J* = 1.8 Hz, 4H, Cp*H*), 4.08–4.06 (m, 4H, Cp*H*), 3.84–3.78 (m, 2H, CH_2), 3.78–3.70 (m, 6H, 2 x OCH_3), 3.67 (dtd, *J* = 8.0, 6.1, 1.8 Hz, 2H, CH_2), 3.64–3.52 (m, 2H, NC*H*(CH_3)$_2$), 3.44–3.39 (m, 5H, 2 x FcCH_2, 5'-*H*), 3.34 (dd, *J* = 10.7, 5.0 Hz, 1H, 5'-*H*), 2.54 (td, *J* = 5.8, 0.8 Hz, 2H, CH_2), 2.49 (ddd, *J* = 13.7, 6.2, 3.5 Hz, 1H, 2'-*H*), 2.43–2.33 (m, 1H, 2'-*H*), 2.30 (t, *J* = 6.7 Hz, 2H, CH_2), 2.00–1.98 (m, 2H, CH_2), 1.20–1.05 (m, 12H, 4 x CH_3). **^{13}C NMR** (126 MHz, CD_3CN) δ 161.9 (*C*O), 159.4 (Ar*C*), 150.6 (Ar*C*), 145.9 (Ar*C*), 139.6 (Ar*C*), 136.7 (Ar*C*), 136.4 (Ar*C*), 131.0 (Ar*C*), 129.0 (Ar*C*), 128.8 (Ar*C*), 127.7 (Ar*C*), 123.4 (Ar*C*), 119.3 (*C*N), 114.1 (Ar*C*), 87.4 (O*C*Ar$_3$), 86.5 (1'-*C*), 86.3 (4'-*C*), 78.9 (Cp*C*), 74.3 (d, *J* = 16.9 Hz, 3'-*C*), 73.5 (Cp*C*H), 71.5 (Cp*C*H), 71.3 (Cp*C*H), 69.4 (Cp*C*), 68.8 (Cp*C*H), 64.1 (5'-*C*), 59.5 (d, *J* = 18.9 Hz, PO*C*H$_2$), 55.9 (2 x O*C*H$_3$), 52.5 (Fc*C*H$_2$), 49.7 (*C*H$_2$), 48.8 (*C*H$_2$), 44.0 (d, *J* = 12.4 Hz, N*C*H(CH_3)$_2$), 42.0 (2'-*C*), 28.7 (*C*H$_2$), 24.9–24.8 (m, NCH(*C*H$_3$)$_2$), 20.9 (d, *J* = 9.5 Hz, *C*H$_2$CN). **^{31}P NMR** (202 MHz, CD_3CN) δ 148.14, 148.12. **HRMS** (ESI) $C_{66}H_{74}Br_2Fe_2N_8O_8P$ [M+H]$^+$: *m/z* 1409.2422, found 1409.2437.

2-cyanoethyl *N,N*-diisopropyl-(3'-*O*-(5-(1-(3-(bis((1',2',3',4',5'-pentamethyl) ferrocenylmethyl)amino)propyl)-1*H*-1,2,3-triazol-4-yl)-5'-*O*-(bis(4-methoxyphenyl)phenylmethyl)-2'-deoxyuridinyl))-phosphoramidite 111



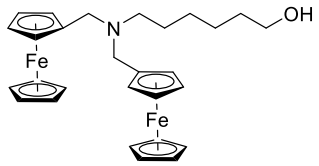
General procedure D was followed using **107** (1.191 g, 1.0 mmol) to yield the title compound as a mixture of diastereomers, as a yellow foam (0.988 g, 53%).

MP 120–122 °C. **IR** ν (cm^{-1}) 3091, 2950, 2247, 1695. **^{1}H NMR** (500 MHz, CD_3CN) δ 8.39 (s, 1H, Ar*H*), 8.04 (s, Ar*H*), 7.47–7.40 (m, 2H, Ar*H*), 7.37–7.29 (m, 4H, Ar*H*), 7.26–7.20 (m, 2H, Ar*H*), 7.18–7.12 (m, 1H, Ar*H*), 6.84–6.80 (m, 4H, Ar*H*), 6.23 (dt, *J* = 7.3, 6.2 Hz, 1H, 1'-*H*), 4.49 (dddt, *J* = 13.3, 9.9, 6.8, 3.4 Hz, 1H, 3'-*H*), 4.21 (td, *J* = 7.0, 2.6 Hz, 2H, CH_2), 4.14 (q, 4.1 Hz, 1H, 4'-*H*), 3.83–3.74 (m, 2H, CH_2), 3.74–3.71 (m, 6H, 2 x OCH_3), 3.68–3.63 (m, 2H, 2 x NC*H*), 3.62 (d, *J* = 1.8 Hz, 4H, Cp*H*), 3.55 (t, *J* = 1.8 Hz, 4H, Cp*H*), 3.41–3.29 (m, 2H, 5'-*H*), 3.19 (s, 4H, FcCH_2), 2.68–2.49 (m, 2H, CH_2), 2.46 (ddd, *J* = 13.7, 6.2, 3.4 Hz, 1H, 2'-*H*), 2.35 (dtd, *J* = 13.7, 6.2, 2.4 Hz, 1H, 2'-*H*), 2.19 (t, *J* = 6.8 Hz, 2H, CH_2), 1.88 (p, *J* = 6.8 Hz, 2H, CH_2), 1.81 (s, 30H, 10 x FcCH_3), 1.19–1.03 (m, 12H, 4 x CH_3) **^{13}C NMR** (126 MHz, CD_3CN) δ 162.0 (*C*O), 159.6 (Ar*C*), 150.6 (Ar*C*), 146.0 (Ar*C*), 139.6 (Ar*C*), 139.6 (Ar*C*), 136.7 (Ar*C*), 136.7 (Ar*C*), 136.7 (Ar*C*), 136.3 (Ar*C*), 131.1 (Ar*C*), 131.1 (Ar*C*), 131.0 (Ar*C*), 128.9 (Ar*C*), 128.9 (Ar*C*), 128.8 (Ar*C*), 127.7 (Ar*C*), 123.4 (Ar*C*), 119.43 (*C*N), 114.1 (Ar*C*), 87.3 (O*C*Ar$_3$), 86.5 (1'-*C*), 86.3 (d, *J* = 4.3 Hz, 4'-*C*) 83.1 (Cp*C*), 80.8 (Cp*C*), 74.5 (d, *J* = 17.6 Hz, 3'-*C*), 73.43 (Cp*C*H), 72.9 (Cp*C*), 64.4 (5'-*C*), 59.6 (d, *J* = 3.0 Hz, PO*C*H$_2$), 55.9 (O*C*H$_3$), 55.9 (O*C*H$_3$), 52.3 (Fc*C*H$_2$), 49.5 (*C*H$_2$), 48.9 (*C*H$_2$), 44.0 (d, *J* = 12.4 Hz, N*C*H(CH$_3$)$_2$), 40.3 (d, *J* = 3.8 Hz, 2'-*C*), 28.8 (*C*H$_2$), 25.0–24.7 (m, NCH(*C*H$_3$)$_2$), 21.0 (d, *J* = 7.2 Hz, *C*H$_2$CN), 11.4 (10 x *C*H$_3$). **^{31}P NMR** (202 MHz, CD_3CN) δ 148.16, 148.12. **HRMS** (ESI) calculated for $C_{76}H_{95}Fe_2N_8O_8PNa$ [M+Na]$^+$: *m/z* 1413.5607, found 1413.5611.

General procedure E for reductive amination of ferrocene carboxaldehydes

To a stirred solution of the ferrocenecarboxaldehyde (1 equiv.) and 6-aminohexan-1-ol (0.6 equiv.) in anhydrous THF (10 mL/mmol) under nitrogen was added sodium tris(acetoxy)borohydride (2.5 equiv.) portionwise over 5 minutes. Once addition was complete the red slurry was allowed to stir at rt overnight. The yellow solution was poured into $NaHCO_{3(sat.)(aq.)}$ (10 mL/mmol) and the aqueous layer was extracted with Et_2O (3 x 50 mL/mmol). The combined organics were dried over $MgSO_4$, filtered, and dried *in vacuo*. The crude oil was purified *via* silica gel chromatography, eluting with petrol/EtOAc/Et_3N (70:30:2) to yield the desired compounds.

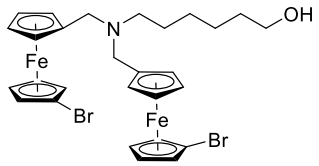
6-(bis(ferrocenylmethyl)amino)hexan-1-ol 71[155]



Using general procedure E with ferrocenecarboxaldehyde (2.14 g, 10 mmol) and 6-aminohexan-1-ol (585 mg, 5 mmol) gave the title compound as an orange solid (2.10 g, 82%).

MP 65–68 °C. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.21 (t, *J* = 1.8 Hz, 4H, Cp*H*), 4.01 (t, *J* = 1.8 Hz, 4H, Cp*H*), 4.00 (s, 10H, Cp*H*), 3.51 (s, 4H, FcC*H2*), 3.39 (t, *J* = 7.1 Hz, 2H, CH_2), 2.44 (t, *J* = 7.1Hz, 2H, CH_2), 1.51 (p, *J* = 7.1 Hz, 2H, CH_2), 1.39 (p, *J* = 6.5 Hz, 2H, CH_2), 1.33–1.21 (m, 4H, 2 x CH_2). **^{13}C NMR** (126 MHz, C_6D_6) δ 84.7 (Cp*C*), 70.4 (Cp*C*H), 68.8 (Cp*C*H), 68.0 (Cp*C*H), 62.6 (*C*H_2), 53.2 (Fc*C*H_2), 52.2 (*C*H_2), 33.1 (*C*H_2), 27.6 (*C*H_2), 27.3 (*C*H_2), 25.9 (*C*H_2).

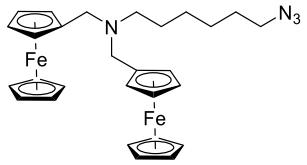
6-(bis((1'-bromo)ferrocenylmethyl)amino)hexan-1-ol 73[155]



Using general procedure E with **46** (0.90 g, 3.1 mmol) to yield the title compound as an orange solid (0.65 g, 32%).

MP 68–69 °C. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.18 (s, 8H, Cp*H*), 4.00 (s, 4H, Cp*H*), 3.72 (t, *J* = 1.5 Hz, 4H, Cp*H*), 3.51 (s, 4H, FcC*H2*), 3.39 (t, *J* = 6.5 Hz, 2H, CH_2), 2.41 (t, *J* = 7.1 Hz, 2H, CH_2), 1.58–1.49 (m, 2H, CH_2), 1.44–1.37 (m, 2H, CH_2), 1.31–1.24 (m, 4H, CH_2), 1.18 (s, 1H, O*H*). **^{13}C NMR** (126 MHz, C_6D_6) δ 86.2 (Cp*C*), 78.7 (Cp*C*), 72.9 (Cp*C*H), 71.0 (Cp*C*H), 70.6 (Cp*C*H), 68.8 (*C*H_2), 67.8 (Cp*C*H), 62.5 (*C*H_2), 52.2 (Fc*C*H_2), 33.1 (*C*H_2), 27.6 (*C*H_2), 27.4 (*C*H_2), 25.8 (*C*H_2).

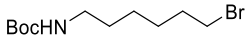
6-azido-*N,N*-bis(ferrocenylmethyl)hexan-1-amine 78



To a stirred solution of 6-(bis(ferrocenylmethyl)amino)hexan-1-ol (3.00 g, 5.9 mmol) and CBr_4 (2.15 g, 6.5 mmol) in anhydrous CH_2Cl_2 (24 mL) at 0 °C was added triphenylphosphine (1.70 g, 6.5 mmol) portion wise, and after addition was complete the solution was allowed to warm to rt. The solution was stirred for 4 hours, then concentrated *in vacuo* to a dark oil. The oil was suspended in hexane (100 mL) and stirred vigorously for 30 minutes. The solution was filtered and the organics were concentrated and filtered through silica, washing with petrol/EtOAc (1:1) until the washings ran clear. The solution was concentrated to give an intermediate brown oil. The oil was dissolved in acetone (24 mL) and water (8 mL), and treated with NaN_3 (1.03 g, 15. 9 mmol) then allowed to stir at 70 °C for 18 hours. The reaction was concentrated *in vacuo* and the oily residue was partitioned between EtOAc (30 mL) and water (10 mL), and the aqueous layer was extracted with EtOAc (3 x 20 mL). The combined organics were washed with $brine_{(sat.)}$ (20 mL), dried over $MgSO_4$, filtered, and concentrated *in vacuo*, which was purified *via* flash chromatography, eluting with petrol/EtOAc/Et_3N (80:20:5) to yield a yellow oil (1.80 g, 57%).

IR ν (cm^{-1}) 3094, 2929, 2857, 2802, 2091. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.14 (t, *J* = 1.7 Hz, 4H, Cp*H*), 3.97 (t, *J* = 1.7 Hz, 4H, Cp*H*), 3.96 (s, 10H, Cp*H*), 3.43 (s, 4H, FcC*H*$_2$), 2.65 (t, *J* = 7.0 Hz, 2H, C*H*$_2$), 2.32 (t, *J* = 7.0 Hz, 2H, C*H*$_2$), 1.35 (p, *J* = 7.0 Hz, 2H, C*H*$_2$), 1.16 (p, *J* = 7.0 Hz, 2H, C*H*$_2$), 1.11–1.04 (m, 2H, C*H*$_2$), 0.99 (m, 2H, C*H*$_2$). **^{13}C NMR** (126 MHz, C_6D_6) δ 83.9 (CpC), 70.1 (Cp*C*H), 68.6 (Cp*C*H), 67.7 (Cp*C*H), 52.9 (*C*H$_2$), 51.6 (Fc*C*H$_2$), 50.9 (*C*H$_2$), 28.6 (*C*H$_2$), 27.8 (*C*H$_2$), 26.6 (*C*H$_2$), 26.3 (*C*H$_2$). **HRMS** (ESI) calculated for $C_{28}H_{35}Fe_2N_4$ $[M+H]^+$: *m/z* 539.1556, found: 539.1534.

tert-butyl (6-bromohexyl)carbamate 82[198]

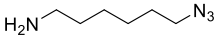


To a stirred solution of 6-aminohexan-1-ol (468 mg, 4 mmol) in anhydrous THF (3 mL) was added di-*tert*-butyl dicarbonate (872 mg, 4 mmol) in one portion and the clear solution was allowed to stir for 3 hours. After this time, the reaction was concentrated *in vacuo* to yield a colourless oil.

The oil was dissolved in anhydrous CH_2Cl_2 (3 mL) and treated with tetrabromomethane (1.46 g, 4.4 mmol) and cooled to 0 °C. The solution was treated with triphenylphosphine (1.15 g, 4.4 mmol) portionwise over 10 minutes. The reaction was allowed to warm to rt and stirred for 30 minutes. The solution was concentrated *in vacuo* to give a thick oil, which was suspended in hexane (50 mL) and stirred vigorously for 30 minutes. The solution was filtered, and the slurry was suspended further with hexane (50 mL) and filtered. The solution was concentrated *in vacuo*, filtered through silica (petrol/EtOAc 1:1) and concentrated *in vacuo* to yield the desired compound as pale-yellow oil (817 mg, 73%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 4.62 (s, 1H, N*H*), 3.58 (t, *J* = 6.6 Hz, 2H, CH_2), 3.07 (q, *J* = 6.6 Hz, 2H, CH_2), 1.52 (p, *J* = 6.6 Hz, 2H, CH_2), 1.48–1.42 (m, 2H, CH_2), 1.40 (s, 9H, CH_3), 1.38–1.26 (m, 4H, 2 x CH_2). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 156.2 (*C*O), 79.1(C $^{\circ 4}$), 62.6 (*C*H_2), 40.5 (*C*H_2), 31.3 (*C*H_2), 30.1 (*C*H_2), 28.5 (3 x *C*H_3), 26.5 (*C*H_2), 25.4 (*C*H_2).

6-azidohexan-1-amine 84[199]



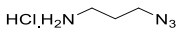
To a stirred solution of 6-aminohexan-1-ol (468 mg, 4 mmol) in anhydrous THF (3 mL) was added di-*tert*-butyl dicarbonate (872 mg, 4 mmol) in one portion and the clear solution was allowed to stir for 3 hours. After this time, the reaction was concentrated *in vacuo* to yield a colourless oil.

The oil was dissolved in toluene (20 mL) and treated with methanesulfonyl chloride (308 µL, 4 mmol) and cooled to 0 °C. The reaction was treated with triethylamine (561 µL, 4 mmol) dropwise, and after addition was complete, the solution was allowed to warm to rt over 20 minutes. The suspension was then treated with tetrabutylammonium bromide (128 mg, 0.4 mmol), NaN_3 (2.21 g, 34 mmol) and H_2O (10 mL), and heated to 70 °C overnight. The reaction was cooled to rt and the aqueous layer was extracted with EtOAc (3 x 100 mL). The combined organics were washed with $brine_{(sat.)}$ (50 mL), dried over $MgSO_4$, filtered, and concentrated *in vacuo* to yield a crude oil.

The oil was dissolved in CH_2Cl_2 (10 mL) and treated with trifluoroacetic acid (918 µL, 12 mmol) dropwise and allowed to stir at rt overnight. The reaction was concentrated *in vacuo* and partitioned between CH_2Cl_2 (50 mL) and $NaHCO_{3(sat.)(aq.)}$ (50 mL). The aqueous layer was extracted with CH_2Cl_2 (3 x 25 mL), and the combined organics were washed with $brine_{(sat.)}$ (20 mL), dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude oil was filtered through silica (petrol/EtOAc 1:1) and concentrated *in vacuo* to yield the desired compound as a colourless oil (510 mg, 90%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 5.31 (s, 2H, N*H*$_2$), 3.19 (d, *J* = 6.9 Hz, 2H, C*H*$_2$), 2.72 (t, *J* = 7.9 Hz, C*H*$_2$), 1.58–1.44 (m, 4H, 2 x C*H*$_2$), 1.34–1.20 (m, 4H, 2 x C*H*$_2$). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 51.2 (CH_2), 40.2 (*C*H_2), 28.5 (*C*H_2), 26.1 (*C*H_2), 25.9 (*C*H_2).

3-azidopropan-1-amine hydrochloride 87[200]



To a stirred solution of 3-aminopropan-1-ol (3.0 mL, 40 mmol) in CH_2Cl_2 (40 mL) at 0 °C was added Boc_2O (9.59 g, 44 mmol), and the solution was allowed to stir at rt overnight. The solution was poured into $NaHCO_{3(sat.)(aq.)}$ (100 mL) and the aqueous layer was extracted with CH_2Cl_2 (3 x 100 mL). The combined organics were washed with $brine_{(sat.)}$ (50 mL), dried over $MgSO_4$, filtered, and concentrated *in vacuo*.

The crude oil was dissolved in toluene (100 mL), cooled to 0 °C, and treated with methanesulfonyl chloride (3.35 mL, 40 mmol), then triethylamine (5.6 mL, 40 mmol) dropwise, stirring vigorously. After addition was complete, the solution was allowed to warm to rt over 20 minutes, then suspension was treated with tetrabutylammonium bromide (1.28 g, 4 mmol), NaN_3 (10.4 g, 160 mmol) and H_2O (20 mL), and heated to 70 °C overnight. The reaction was cooled to rt and the aqueous layer was extracted with EtOAc (3 x 100 mL). The combined organics were washed with $brine_{(sat.)}$ (50 mL), dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude oil was filtered through silica (petrol/EtOAc 1:1) and concentrated.

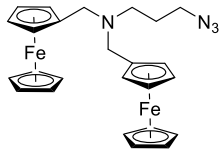
The crude oil was dissolved in 2M HCl in Et_2O (40 mL, 80 mmol) and allowed to stir in a sealed flask for 24 h. The reaction was concentrated *in vacuo* and the gum was re-suspended in Et_2O. The precipitate was filtered without further purification, yielding the title compound as a white solid (3.44 g, 63%).

^{1}H NMR (500 MHz, D_2O) δ 3.57–3.49 (m, 2H, CH_2), 3.18–3.08 (m, 2H, CH_2), 2.02–1.93 (m, 2H, CH_2). **^{13}C NMR** (126 MHz, D_2O) δ 48.2 (*C*H_2), 37.3 (*C*H_2), 26.0 (*C*H_2).

General procedure F for the synthesis of di-ferrocenyl azides

To a stirred solution of the corresponding ferrocenecarboxaldehyde (1 equiv.) and 3-azidopropan-1-amine hydrochloride, **87** (0.6 equiv.) in anhydrous THF (3 mL/mmol) at 0 °C was added sodium tris(acetoxy)borohydride (3.5 equiv.) portionwise and the slurry was allowed to stir at rt overnight. The solution was quenched $NaHCO_{3(sat.)(aq.)}$ and extracted with Et_2O (3 x 50 mL/mmol). The combined organics were washed with $brine_{(sat.)}$ (25 mL), dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude oils were purified *via* silica gel chromatography to yield the desired compounds.

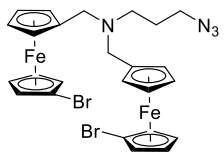
3-azido-*N,N*-bis(ferrocenylmethyl)propan-1-amine 88



General procedure F was followed using ferrocenecarboxaldehyde (1.28 g, 6.0 mmol), **87** (500 mg, 3.6 mmol), sodium tris(acetoxy)borohydride (4.45 g, 21 mmol), and THF (20 mL), eluting with petrol/EtOAc/Et_3N (80:20:5) to yield the title compound as an orange oil (1.06 g, 72%)

IR ν (cm^{-1}) 3092, 2948, 2809, 2093. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.11 (t, *J* = 1.8 Hz, 4H, Cp*H*), 3.99 (t, *J* = 1.8 Hz, 4H, Cp*H*), 3.98 (s, 10H, Cp*H*), 3.38 (s, 4H, FcCH_2), 2.90 (t, *J* = 6.7 Hz, 1H, CH_2), 2.32 (t, *J* = 6.7 Hz, 1H, CH_2), 1.43 (p, *J* = 6.7 Hz, 1H, CH_2). **^{13}C NMR** (126 MHz, C_6D_6) δ 84.3 (Cp*C*), 70.3 (Cp*C*H), 68.9 (Cp*C*H), 68.2 (Cp*C*H), 53.2 (CH_2), 49.5 (CH_2), 49.3 (CH_2), 27.2 (CH_2). **HRMS** (ESI) calculated for $C_{25}H_{29}FeN_4$ $[M+H]^+$: *m/z* 497.1086, found 497.1173.

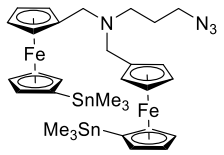
3-azido-*N,N*-bis((1'-bromo)ferrocenylmethyl)propan-1-amine 98



General procedure F was followed using **46** (1.88 g, 6.4 mmol), **87** (522 mg, 3.8 mmol), sodium tris(acetoxy)borohydride (4.74 g, 22.4 mmol), and THF (20 mL), eluting with petrol/EtOAc/Et_3N (80:20:5) to yield the title compound as an orange oil (1.55 g, 74%).

IR ν (cm^{-1}) 3087, 2940, 2812, 2090. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.16 (t, *J* = 1.9 Hz, 4H, Cp*H*), 4.08 (t, *J* = 1.9 Hz, 4H, Cp*H*), 3.99 (d, *J* = 1.9 Hz, 4H, Cp*H*), 3.71 (t, *J* = 1.9 Hz, 4H, Cp*H*), 3.38 (s, 4H, FcC*H*$_2$), 2.90 (t, *J* = 6.6 Hz, 2H, C*H*$_2$), 2.29 (t, *J* = 6.6 Hz, 2H, C*H*$_2$), 1.45 (p, *J* = 6.6 Hz, 2H, C*H*$_2$). **^{13}C NMR** (126 MHz, C_6D_6) δ 85.9 (Cp*C*), 78.7 (Cp*C*), 72.9 (Cp*C*H), 71.0 (Cp*C*H), 70.7 (Cp*C*H), 67.9 (Cp*C*H), 52.2 (Fc*C*H$_2$), 49.5 (*C*H$_2$), 49.5 (*C*H$_2$), 27.2 (*C*H$_2$). **HRMS** (ESI) calculated for $C_{25}H_{27}Br_2Fe_2N_4$ $[M+H]^+$: *m/z* 654.9275, found 654.9326.

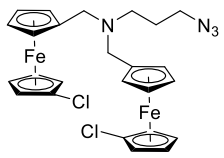
3-azido-*N,N*-bis((1'-trimethylstannyl)ferrocenylmethyl)propan-1-amine 100



General procedure F was followed **97** (1.13 g, 3 mmol), **87** (244 mg, 1.8 mmol), sodium tris(acetoxy)borohydride (2.22 g, 10.5 mmol), and THF (10 mL), eluting with petrol/EtOAc/Et_3N (80:20:5) to yield the title compound as an orange oil (749 mg, 60%).

IR ν (cm^{-1}) 3081, 2930, 2834, 2093. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.20 (t, *J* = 1.8 Hz, 4H, Cp*H*), 4.13 (t, *J* = 1.8 Hz, 4H, Cp*H*), 4.02 (t, *J* = 1.8 Hz, 4H, Cp*H*), 3.93 (t, *J* = 1.8 Hz, 4H, Cp*H*), 3.42 (s, 4H, FcC*H*$_2$), 2.92 (t, *J* = 6.7 Hz, 2H, C*H*$_2$), 2.33 (t, *J* = 6.7 Hz, 2H, C*H*$_2$), 1.47 (p, *J* = 6.7 Hz, 2H, C*H*$_2$), 0.27 (s, 12H, 6 x SnC*H*$_3$). **^{13}C NMR** (126 MHz, C_6D_6) δ 84.1 (Cp*C*), 75.0 (Cp*C*H), 71.8 (Cp*C*H), 70.4 (Cp*C*H), 69.1 (Cp*C*), 68.2 (Cp*C*H), 53.4 (Fc*C*H$_2$), 49.5 (*C*H$_2$), 49.4(*C*H$_2$), 27.2 (*C*H$_2$), -8.6 (6 x Sn*C*H$_3$). **HRMS** (ESI) calculated for $C_{31}H_{45}Fe_2N_4Sn_2$ $[M+H]^+$: *m/z* 825.0383, found 825.0355.

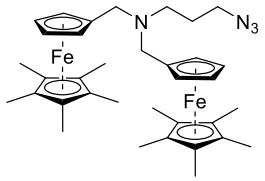
3-azido-*N,N*-bis(chloroferrocenylmethyl)propan-1-amine 101



General procedure F was followed using **91** (2.22 g, 8.9 mmol), **87** (726 mg, 5.34 mmol), sodium tris(acetoxy)borohydride (6.59 g, 31.1 mmol), and THF (35 mL), eluting with petrol/EtOAc/Et_3N (80:20:5) to yield the title compound as an orange oil (891 mg, 18%).

IR ν (cm^{-1}) 3089, 2975, 2882, 2089. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.13 (t, *J* = 1.9 Hz, 4H (Cp*H*), 4.11 (t, *J* = 1.9 Hz, 4H, Cp*H*), 3.99 (t, *J* = 1.9 Hz, 4H, Cp*H*), 3.67 (t, *J* = 1.9 Hz, 4H, Cp*H*), 3.39 (s, 4H, FcCH_2), 2.89 (t, *J* = 6.7 Hz, 2H, CH_2), 2.30 (t, *J* = 6.7 Hz, 2H, CH_2), 1.44 (p, *J* = 6.7 Hz, 2H, CH_2). **^{13}C NMR** (126 MHz, C_6D_6) δ 85.8 (Cp*C*), 72.5 (Cp*C*H), 70.4 (Cp*C*H), 68.7 (Cp*C*H), 67.6 (Cp*C*), 66.9 (Cp*C*H), 52.3 (Fc*C*H_2), 49.5 (*C*H_2), 49.5 (*C*H_2), 27.2 (*C*H_2). **HRMS** (ESI) calculated for $C_{25}H_{26}Cl_2Fe_2N_4Na$ $[M+Na]^+$: *m/z* 587.0127, found 587.0126.

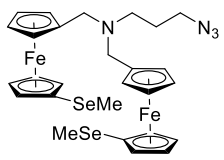
3-azido-*N,N*-bis((1',2',3',4',5'-pentamethyl)ferrocenylmethyl)propan-1-amine 102



General procedure F was followed using **95** (1.80 g, 6.3 mmol), **87** (514 g, 3.8 mmol), sodium tris(acetoxy)borohydride (4.67 g, 22.1 mmol), and THF (20 mL) petrol/EtOAc/Et_3N (90:10:5) to yield the title compound as an orange oil (875 mg, 43%).

IR ν (cm^{-1}) 3086, 2981, 2941, 2875, 2092. **^{1}H NMR** (300 MHz, C_6D_6) δ 3.66 (t, *J* = 1.8 Hz, 4H, Cp*H*), 3.62 (t, *J* = 1.8 Hz, 4H, Cp*H*), 3.34 (s, 4H, FcCH_2), 2.93 (t, *J* = 6.7 Hz, 2H, CH_2), 2.37 (t, *J* = 6.7 Hz, 2H, CH_2), 1.80 (s, 30H, 10 x CH_3), 1.52 (p, *J* = 6.7 Hz, 2H, CH_2). **^{13}C NMR** (75 MHz, C_6D6) δ 83.2 (Cp*C*), 80.4 (Cp*C*), 73.1 (Cp*C*H), 72.7 (Cp*C*H), 52.5 (Fc*C*H_2), 49.9 (*C*H_2), 49.8 (*C*H_2), 27.7 (*C*H_2), 11.7 (10 X *C*H_3). **HRMS** (ESI) calculated for $C_{35}H_{49}Fe_2N_4$ $[M+H]^+$: *m/z* 637.2651, found 637.2702.

3-azido-*N,N*-bis((1'-methylseleno)ferrocenylmethyl)propan-1-amine 105



To a stirred solution of **47** (561 mg, 1.6 mmol) in CH_2Cl_2 (5 mL) and H_2O (2.5 mL) was added *p*-toluenesulfonic acid (631 mg, 1.8 mmol) and the solution was allowed to stir at rt for 30 minutes. The red solution was quenched with $NaHCO_{3(sat.)(aq.)}$ (20 mL) and the aqueous layer was extracted with CH_2Cl_2 (3 x 10 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated *in vacuo* to give the crude aldehyde.

The crude aldehyde was dissolved in anhydrous THF (5 mL) and treated with **87** (130 mg, 0.96 mmol) and allowed to stir at rt for 10 minutes. To the solution was added sodium tris(acetoxy)borohydride (1.19 g, 5.6 mmol) portionwise and the reaction was allowed to stir at rt overnight. The reaction was quenched with $NaHCO_{3(sat.)(aq.)}$ (30 mL) and the aqueous layer was extracted with Et_2O (3 x 20 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated *in vacuo*, then the crude oil was purified *via* silica gel chromatography eluting with petrol/EtOAc/Et_3N (80:20:5) to yield the title compound as an orange oil (380 mg, 70%).

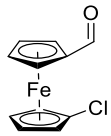
IR ν (cm^{-1}) 3087, 2926, 2815, 2089. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.18 (t, *J* = 1.8 Hz, 4H, Cp*H*), 4.14 (t, *J* = 1.8 Hz, 4H, Cp*H*), 4.04 (t, *J* = 1.8 Hz, 4H, Cp*H*), 3.92 (t, *J* = 1.8 Hz, 4H, Cp*H*), 3.46 (s, 4H, FcC*H*$_2$), 2.91 (t, *J* = 6.7 Hz, 2H, C*H*$_2$), 2.30 (t, *J* = 6.7 Hz, 2H, C*H*$_2$), 1.89 (s, 6H, C*H*$_3$), 1.47 (p, *J* = 6.7 Hz, 2H, C*H*$_2$). **^{13}C NMR** (126 MHz, C_6D_6) δ 84.8 (Cp*C*), 74.4 (Cp*C*H), 71.4 (Cp*C*H), 69.7 (Cp*C*H), 69.2 (Cp*C*H), 68.5 (Cp*C*), 52.3 (Fc*C*H$_2$), 49.1 (*C*H$_2$), 49.0 (*C*H$_2$), 26.8 (*C*H$_2$), 9.1 (*C*H$_3$). **HRMS** (ESI) calculated for $C_{27}H_{33}Fe_2N_4Se_2$ $[M+H]^+$: *m/z* 684.9729 found 684.9761.

Chloroferrocene 90[162]

To a stirred solution of ferrocene (6.00 g, 32.2 mmol) and potassium *tert*-butoxide (448 mg, 4.0 mmol) in anhydrous THF (280 mL) at -78 °C was added *t*-BuLi (1.7 M in pentane) (38 mL, 64.4 mmol) dropwise over 20 minutes *via* pressure equalising funnel, forming an orange precipitate, and the suspension was allowed to stir for 90 minutes at -78 °C. After this time the suspension was treated with hexachloroethane (11.44 g, 48.5 mmol) in anhydrous THF (20 mL) dropwise, and once addition was complete was allowed to stir for 30 minutes at -78 °C. The dark orange solution was warmed to rt and quenched by the addition of H_2O (250 mL), and the aqueous layer was extracted with EtOAc (3 x 100 mL). The combined organics were concentrated to 200 mL, and washed with 0.2 M $FeCl_{3(aq.)}$ (4 x 100 mL), H_2O (3 x 100 mL), and $brine_{(sat.)}$ (100 mL). The organic layer was dried over $MgSO_4$, filtered, and concentrated *in vacuo* to yield the title compound as an amorphous orange solid (4.22 g, 60%).

MP 63–64 ° C. **^{1}H NMR** (500 MHz, C_6D_6) δ 4.20 (t, *J* = 1.9 Hz, 2H, Cp*H*), 4.00 (s, 5H, Cp*H*), 3.68 (t, *J* = 1.9 Hz, 2H, Cp*H*). **^{13}C NMR** (126 MHz, C_6D_6) δ 92.3 (Cp*C*), 70.6 (Cp*C*H), 68.2 (Cp*C*H), 66.3 (Cp*C*H).

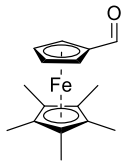
Chloroferrocenecarboxaldehyde 91[201]



To a solution of phosphorous oxychloride (5.06 mL, 54.4 mmol) and *N*-methylformanilide (10.25 mL, 83.2 mmol) was added **90** (7.03 g, 32.0 mmol). The slurry was stirred at rt for 1 hour, then heated to 70 °C for 2 hours. The slurry was cooled to rt and quenched by the addition of NaOAc (32 g) in H_2O (200 mL) and allowed to stir overnight. The biphasic mixture was filtered through celite and washed with $CHCl_3$ (2.5 L) until the washings ran clear. The aqueous layer was extracted with $CHCl_3$ (3 x 50 mL), and the combined organics were dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude red oil was purified twice *via* flash chromatography, eluting with petroleum ether/ EtOAc (9:1) to yield the title compound as a 4:1 ratio of inseparable 1,2-, and 1,1'- regioisomers, as a red solid (2.22 g, 27%).

MP 56–58 °C. **^{1}H NMR** (500 MHz, C_6D_6) δ 9.78 (s, 1H, C*H*O), 4.50 (s, 2H, Cp*H*), 4.07 (s, 2H, Cp*H*), 4.05 (s, 2H, Cp*H*), 3.56 (s, 2H, Cp*H*). **^{13}C NMR** (126 MHz, C_6D_6) δ 191.8 (*C*HO), 92.9 (Cp*C*), 74.7 (Cp*C*H), 71.5 (Cp*C*H), 71.1 (Cp*C*H), 68.7 (Cp*C*), 67.3 (Cp*C*H).

(1',2',3',4',5'-pentamethyl)ferrocenecarboxaldehyde 95[202]



To a round bottom flask under argon was added freshly cracked cyclopentadiene (1.66 mL, 19.8 mmol) and treated with 2.5 M *n*-butyllithium (8.72 mL,21.8 mmol) dropwise at -78 °C and was allowed to warm to rt and stirred for 1 hour. The white slurry was treated with methyl formate (1.81 mL, 29.7 mmol) and allowed to stir at rt for 4 hours to give a deep red colour.

2 hours after initial reaction, to a stirred solution of 1,2,3,4,5-pentamethylcyclopentadiene (3.4 mL, 21.8 mmol) in THF (100 mL) under argon at -78 °C was added 2.5 M *n*-butyllithium (8.72 mL, 21.8 mmol) dropwise and was allowed to warm to rt and stirred for 2 hours.

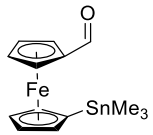
3 hours after initial reaction, a suspension of $FeCl_2$ (2.5 g, 19.8 mmol) in THF (100 mL) was vigorously stirred in the dark for 1 hour.

4 hours after initial reaction, to the iron chloride suspension was added the LiCp* *via* cannula transfer with argon pressure and the mixture was allowed to stir at rt for 1 hour. 5 hours after initial reaction, the LiCp was added to the LiCp* + $FeCl_2$ mixture *via* cannula transfer and was allowed to stir at rt for 90 minutes.

The solution was concentrated *in vacuo* and redissolved in 1:1 Et_2O/Hexane. The solution was filtered through celite until the washings ran clear. The solution was concentrated *in vacuo* and purified *via* silica gel chromatography, eluting with petroleum ether/EtOAc (9:1) to yield the title compound as a red solid (1.89 g, 34%).

MP 60–62 °C. **^{1}H NMR** (300 MHz, C_6D_6) δ 9.73 (s, 1H, C*H*O), 4.11 (t, *J* = 1.8 Hz, 2H, Cp*H*), 3.78 (t, *J* = 1.8 Hz, 2H, Cp*H*), 1.62 (s, 15H, 5 x C*H*$_3$). **^{13}C NMR** (75 MHz, C_6D_6) δ 192.5 (*C*O), 81.9 (Cp*C*H), 81.1 (Cp*C*), 77.0 (Cp*C*), 72.1 (Cp*C*H), 10.9 (5 x *C*H_3).

1'-(trimethylstannyl)ferrocenecarboxaldehyde 97



To a stirred solution of **96** (1.55 g, 3.6 mmol) in CH_2Cl_2 (10 mL) and H_2O (5 mL) was added *p*-toluenesulfonic acid (681 mg, 3.96 mmol) and the reaction was allowed to stir for 30 minutes. The reaction was poured into $NaHCO_{3(sat.)(aq.)}$ and the aqueous layer was extracted with CH_2Cl_2 (3 x 10 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude oil was purified *via* silica gel chromatography eluting with petroleum ether/EtOAc (9:1) to yield the title compound as a red oil (1.15 g, 85%).

IR ν (cm^{-1}) 3085, 2978, 2909, 2818, 1679. **^{1}H NMR** (500 MHz, C_6D_6) δ 9.87 (s, 1H, C*H*O), 4.53 (s, 2H, Cp*H*), 4.16 (s, 2H, Cp*H*), 4.12 (s, 2H, Cp*H*), 3.88 (s, 2H, Cp*H*), 0.19 (s, 9H, 3 x C*H*$_3$). **^{13}C NMR** (126 MHz, C_6D_6) δ 191.9 (*C*HO), 80.3 (Cp*C*), 75.7 (Cp*C*H), 72.7 (Cp*C*H), 72.3 (Cp*C*H), 71.1 (Cp*C*), 69.6 (Cp*C*H), -8.9 (3 x *C*H_3). **HRMS** (ESI) calculated for $C_{14}H_{18}FeOSnNa$ $[M+Na]^+$: *m/z* 400.9621, found 400.9672.

References

1 M. Burnworth, S. J. Rowan and C. Weder, *Chem . Eur. J.*, 2007, **13**, 7828–7836.

2 P. A. Emanuel, J. Dang, J. S. Gebhardt, J. Aldrich, E. A. E. Garber, H. Kulaga, P. Stopa, J. J. Valdes and A. Dion-Schultz, *Biosens. Bioelectron.*, 2000, **14**, 751–759.

3 J. Liu and B. Mattiasson, *Water Res.*, 2002, **36**, 3786–3802.

4 B. D. Malhotra and A. Chaubey, *Sens. Actuators B-Chem.*, 2003, **91**, 117–127.

5 E. B. Bahadir and M. K. Sezgintürk, *Anal. Biochem.*, 2015, **478**, 107–120.

6 O. Lazcka, F. Javier, D. Campo and F. Xavier Muñoz, *Biosens. Bioelectron.*, 2007, **22**, 1205–1217.

7 R. Singh, M. Das Mukherjee, G. Sumana, R. K. Gupta, S. Sood and B. D. Malhotra, *Sens. Actuators B-Chem.*, 2014, **197**, 385–404.

8 R. Hnasko, Ed., *ELISA*, Springer New York, 2015, vol. 1318.

9 C.-M. Cheng, A. W. Martinez, J. Gong, C. R. Mace, S. T. Phillips, E. Carrilho, K. A. Mirica and G. M. Whitesides, *Angew. Chem. Int. Ed.*, 2010, **49**, 4771–4774.

10 D. M. Rissin, C. W. Kan, T. G. Campbell, S. C. Howes, D. R. Fournier, L. Song, T. Piech, P. P. Patel, L. Chang, A. J. Rivnak, E. P. Ferrell, J. D. Randall, G. K. Provuncher, D. R. Walt and D. C. Duffy, *Nat. Biotechnol.*, 2010, **28**, 595–599.

11 H. Fernández, F. J. Arévalo, A. M. Granero, S. N. Robledo, C. H. D. Nieto, W. I. Riberi and M. A. Zon, *Chemosensors*, 2017, **5**, 23.

12 G. Wu and M. H. Zaman, *Bull. World Health Organ.*, 2012, **90**, 914–920.

13 A. Baumstark, N. Jendrike, S. Pleus, C. Haug and G. Freckmann, *Diabetes Technol. Ther.*, 2017, **19**, 580–588.

14 A. M. Rompalo, Y.-H. Hsieh, M. T. Hogan, M. Barnes, M. Jett-Goheen, J. S. Huppert and C. A. Gaydos, *Sex. Health*, 2013, **10**, 541–545.

15 M. G. Bloomfield, M. N. D. Balm and T. K. Blackmore, *Pathology*, 2015, **47**, 227–233.

16 S. T. Sadiq, F. Mazzaferri and M. Unemo, *Sex. Transm. Infect.*, 2017, **93**, S65–S68.

17 J. W. F. Law, N. S. Ab Mutalib, K. G. Chan and L. H. Lee, *Front. Microbiol.*, 2015, **6**, 1–15.

18 A. Hadgu, N. Dendukuri and J. Hilden, *Epidemiology*, 2005, **16**, 604–612.

19 E. M. Harding-Esch, A. V Nori, A. Hegazi, M. J. Pond, O. Okolo, A. Nardone, C. M.

Lowndes, P. Hay and S. T. Sadiq, *Sex. Transm. Infect.*, 2017, **0**, 1–6.

20 M. Fakruddin, K. S. Bin Mannan, A. Chowdhury, R. M. Mazumdar, M. N. Hossain, S. Islam and M. A. Chowdhury, *J. Pharm. Bioallied Sci.*, 2013, **5**, 245–52.

21 R. K. Saiki, S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. Erlich and N. Arnheim, *Science*, 1985, **230**, 1350–1354.

22 R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. Erlich, *Science*, 1988, **239**, 487–491.

23 K. Hayashi, M. ORita, Y. Suzuki and T. Sekiya, *Nucleic Acids Res.*, 1989, **17**, 3605.

24 A. Landgraf, B. Reckmann and A. Pingoud, *Anal. Biochem.*, 1991, **193**, 231–235.

25 P. M. Holland, R. D. Abramson, R. Watson and D. H. Gelfand, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 7276–7280.

26 H. D. VanGuilder, K. E. Vrana and W. M. Freeman, *Biotechniques*, 2008, **44**, 619–626.

27 P. S. Bernard and C. T. Wittwer, *Clin. Chem.*, 2002, **48**, 1178–1185.

28 M. J. Espy, J. R. Uhl, L. M. Sloan, S. P. Buckwalter, M. F. Jones, E. A. Vetter, J. D. C. Yao, N. L. Wengenack, J. E. Rosenblatt, F. R. Cockerill III and T. F. Smith, *Clin. Microbiol. Rev.*, 2006, **19**, 165–256.

29 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.

30 J. M. Berg, J. L. Tymoczko, G. J. Gattor Jr. and L. Stryer, *Biochemistry*, W. H. Freeman and Company, New York, 8th edn.

31 S. L. Beaucage and M. H. Caruthers, *Tetrahedron Lett.*, 1981, **22**, 1859–1862.

32 N. D. Sinha, J. Biernat, J. M. Mc Manus and H. Köster, *Nucleic Acids Res.*, 1984, **12**, 4539–4557.

33 J. Hovinen, A. Guzaev, A. Azhayev and H. Lonnberg, *J. Chem. Soc. Perkin Trans. 1*, 1994, 2745–2749.

34 E. Hilario, *Mol. Biotechnol.*, 2004, **28**, 77–80.

35 V. Derbyshire, P. S. Freemont, M. R. Sanderson, L. Beese, J. M. Friedman, C. M. Joyce and T. A. Sterrz, *Science*, 1988, **240**, 199–201.

36 Z. Huang and J. W. Szostak, *Nucleic Acids Res.*, 1996, **24**, 4360–4361.

37 M. D. Challberg and P. T. Englund, *Methods Enzymol.*, 1980, **65**, 39–42.

38 S. Liu, T. Liu and L. Wang, *Chem. Commun.*, 2015, **51**, 176–179.

39 G. Winter and G. G. Brownlee, *Nucleic Acids Res.*, 1978, **5**, 3129–3140.

40 J Lingner and W Keller, *Nucleic Acids Res.*, 1993, **21**, 2917–2920.

41 D. C. Ward, A. Cerami, E. Reich, G. Acs and L. Altwerger, *J. Biol. Chem.*, 1969, **244**, 3243–3250.

42 C.-P. D. Tu and S. N. Cohen, *Gene*, 1980, **10**, 177–183.

43 M-L Fontanel, H Bazin and R Teoule, *Anal. Biochem.*, 1993, **214**, 338–340.

44 B. C. F. Chu, G. M. Wahl and L. E. Orgel, *Nucleic Acids Res.*, 1983, **11**, 6513–6529.

45 A. V. Lebedev and E. Wickstrom, *Perspect. Drug Discov. Des.*, 1996, **4**, 17–40.

46 A. Khvorova and J. K. Watts, *Nat. Biotechnol.*, 2017, **35**, 238–248.

47 H. Krishna and M. H. Caruthers, *J. Am. Chem. Soc.*, 2012, **134**, 11618–11631.

48 G. Chatelain, A. Meyer, F. Morvan, J.-J. Vasseur and C. Chaix, *New J. Chem.*, 2011, **35**, 893.

49 S. H. Weisbrod and A. Marx, *Chem. Commun.* , 2008, 5675–5685.

50 C. J. Yu, Y. Wan, H. Yowanto, J. Li, C. Tao, M. D. James, C. L. Tan, G. F. Blackburn and T. J. Meade, *J. Am. Chem. Soc.*, 2001, **123**, 11155–11161.

51 K. Yamana, Y. Nishijima, T. Ikeda, T. Gokota, H. Ozaki, H. Nakano, O. Sangen and T. Shimidzifl, *Bioconjugate Chem.*, 1990, **1**, 319–324.

52 L. Beielman, A. Karpeisky, J. Matulic-adamic, P. Haeberli, D. Sweedler and N. Usman, *Nucleic Acids Res.*, 1995, **23**, 4434–4442.

53 K. Yamana, T. Mitsui, H. Hayashi and H. Nakano, *Tetrahedron Lett.*, 1997, **38**, 5815–5818.

54 K. Yamana, Y. Ohashi, K. Nunota and H. Nakano, *Tetrahedron*, 1997, **53**, 4265–4270.

55 S. M. Langenegger and R. Häner, *Chem. Commun.*, 2004, 2792–2793.

56 S. M. Langenegger and R. Häner, *ChemBioChem*, 2005, **6**, 2149–2152.

57 B. N. Trawick, T. A. Osiek and J. K. Bashkin, *Bioconjugate Chem.*, 2001, **12**, 900–905.

58 H. Kashida, H. Asanuma and M. Komiyama, *Chem. Commun.* , 2006, **1**, 2768–2770.

59 N. Moran, D. M. Bassani, J.-P. Desvergne, S. Keiper, P. A. S. Lowden, J. S. Vyle and J. H. R. Tucker, *Chem. Commun.* , 2006, 5003–5005.

60 H. V. Nguyen, A. Sallustrau, L. Male, P. J. Thornton and J. H. R. Tucker, *Organometallics*,

2011, **30**, 5284–5290.

61 H. V. Nguyen, Z. Zhao, A. Sallustrau, S. L. Horswell, L. Male, A. Mulas and J. H. R. Tucker, *Chem. Commun.*, 2012, **48**, 12165–12167.

62 W. Xu, K. M. Chan and E. T. Kool, *Nat. Chem.*, 2017, **9**, 1043–1055.

63 W. A. Wlassoff and G. C. King, *Nucleic Acids Res.*, 2002, **30**, 1–7.

64 Y. Hasegawa, T. Takada, M. Nakamura and K. Yamana, *Bioorg. Med. Chem. Lett.*, 2017, **27**, 3555–3557.

65 C. J. Yu, H. Yowanto, Y. Wan, T. J. Meade, Y. Chong, M. Strong, L. H. Donilon, J. F. Kayyem, M. Gozin and G. F. Blackburn, *J. Am. Chem. Soc.*, 2000, **122**, 6767–6768.

66 A. E. Beilstein and M. W. Grinstaff, *Chem. Commun.*, 2000, 509–510.

67 P. W. J. Rigby, M. Dieckmann, C. Rhodes and P. Berg, *J. Mol. Biol.*, 1977, **113**, 237–251.

68 H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew. Chem. Int. Ed.*, 2001, **40**, 2004–2021.

69 R. Huisgen, *Proc. Chem. Soc.*, 1961, **0**, 357–396.

70 C. W. Tornøe, C. Christensen and M. Meldal, *J. Org. Chem.*, 2002, **67**, 3057–3064.

71 V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angew. Chem. Int. Ed.*, 2002, **41**, 2596–2599.

72 A. H. El-Sagheer and T. Brown, *Chem. Soc. Rev.*, 2010, **39**, 1388–1405.

73 J. Gierlich, G. A. Burley, P. M. E. Gramlich, D. M. Hammond and T. Carell, *Org. Lett.*, 2006, **8**, 3639–3642.

74 P. M. E. Gramlich, S. Warncke, J. Gierlich and T. Carell, *Angew. Chem. Int. Ed.*, 2008, **47**, 3442–3444.

75 E.T. Bolton and B. J. McCarthy, *Proc. Natl. Acad. Sci. U. S. A.*, 1962, **48**, 1390–1397.

76 R. Porecha and D. Herschlag, *Methods Enzymol.*, 2013, **530**, 255–279.

77 B. Tavitian, S. Terrazzino, B. Kuhnast, S. Marzabal, O. Stettler, F. Dolle, J.-R. Deverre, A. Jobert, F. Hinnen, B. Bendriem, C. Crouzel and L. Di Giamberardino, *Nat. Med.*, 1998, **4**, 467–470.

78 B. Kuhnast, F. Dollé, F. Vaufrey, F. Hinnen, C. Crouzel and B. Tavitian, *J. Label. Compd. Radiopharm.*, 2000, 43, 837–848.

79 Y. Song, W. Wei and X. Qu, *Adv. Mater.*, 2011, **23**, 4215–4236.

80 M. Chee, R. Yang, E. Hubbell, A. Berno, X. C. Huang, D. Stern, J. Winkler, D. J. Lockhart, M. S. Morris and S. P. A. Fodor, *Science*, 1996, **274**, 610–614.

81 B. Wardle, *Principles and Applications of Photochemistry*, Wiley, Chichester, United Kingdom, 2009, vol. 36.

82 E. A. Meyer, R. K. Castellano and F. Diederich, *Angew. Chem. Int. Ed.*, 2003, **42**, 1210–1250.

83 J.-B. LePecq and C. Paoletti, *J. Mol. Biol.*, 1967, **27**, 87–106.

84 L. T. Jin and J. K. Choi, *Electrophoresis*, 2004, **25**, 2429–2438.

85 R. Rasmussen, T. Morrison, M. Herrmann and C. Wittwer, *Biochemica*, 1998, **2**, 8–11.

86 R. M. Hartshorn and J. K. Barton, *J. Am. Chem. Soc.*, 1992, **114**, 5919–5925.

87 S. Tyagi and F. R. Kramer, *Nat. Biotechnol.*, 1996, **14**, 303–306.

88 D. Whitcombe, J. Theaker, S. P. Guy, T. Brown and S. Little, *Nat. Biotechnol.*, 1999, **17**, 804–807.

89 K. J. Livak, S. J. A. Flood, J. Marmaro, W. Giusti and K. Deetz, *Genome Res.*, 1995, **4**, 357–362.

90 L. G. Lee, C. R. Connell and W. Bloch, *Nucleic Acids Res.*, 1993, **21**, 3761–3766.

91 T. G. Drummond, M. G. Hill and J. K. Barton, *Nat. Biotechnol.*, 2003, **21**, 1192–1199.

92 T. M. Herne and M. J. Tarlov, *J. Am. Chem. Soc.*, 1997, **119**, 8916–8920.

93 P. M. Armistead and H. H. Thorp, *Bioconjugate Chem.*, 2002, **13**, 172–176.

94 H. Takenaka, S. Sato and S. Takenaka, *Electroanalysis*, 2013, **25**, 1827–1830.

95 H. Gaiji, P. Jolly, S. Ustuner, S. Goggins, M. Abderrabba, C. G. Frost and P. Estrela, *Electroanalysis*, 2017, **29**, 917–922.

96 E. L. S. Wong and J. J. Gooding, *Anal. Chem.*, 2006, **78**, 2138–44.

97 E. Paleček and M. A. Hung, *Anal. Biochem.*, 1983, **132**, 236–242.

98 E. Palecek, E. Lukasova, F. Jelen and M. Vojtíšková, *Bioelectrochem. Bioenerg.*, 1981, **8**, 497–506.

99 P. Verspieren, A. W. C. A. Cornelissen, N. T. Thuong, C. Hélène and J. J. Toulmé, *Gene*, 1987, **61**, 307–315.

100 C. Fan, K. W. Plaxco and A. J. Heeger, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 9134–

9137.

101 R. Y. Lai, E. T. Lagally, S.-H. Lee, H. T. Soh, K. W. Plaxco and A. J. Heeger, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 4017–4021.

102 C. E. Immoos, S. J. Lee and M. W. Grinstaff, *J. Am. Chem. Soc.*, 2004, **126**, 10814–10815.

103 Y. Du, B. J. Lim, B. Li, Y. S. Jiang, J. L. Sessler and A. D. Ellington, *Anal. Chem.*, 2014, **86**, 8010–8016.

104 N. Tibanyenda, S. H. De Bruin, C. A. G. Haasnoot, G. A. Van Dek Marel, J. H. Van Boom and C. W. Hilbers, *Eur. J. Biochem*, 1984, **139**, 19–270.

105 Y. Osakada, K. Kawai, M. Fujitsuka and T. Majima, *Nucleic Acids Res.*, 2008, **36**, 5562–5570.

106 J. C. Genereux and J. K. Barton, *Chem. Rev.*, 2010, **110**, 1642–1662.

107 M. Inouye, R. Ikeda, M. Takase, T. Tsuri and J. Chiba, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 11606–10.

108 S. O. Kelley, E. M. Boon, J. K. Barton, N. M. Jackson and M. G. Hill, *Nucleic Acids Res.*, 1999, **27**, 4830–4837.

109 E. M. Boon, D. M. Ceres, T. G. Drummond, M. G. Hill and J. K. Barton, *Nat. Biotechnol.*, 2000, **18**, 1096–1100.

110 A.-E. Navarro, N. Spinelli, C. Moustrou, C. Chaix, B. Mandrand and H. Brisset, *Nucleic Acids Res.*, 2004, **32**, 5310–5319.

111 A. E. Navarro, N. Spinelli, C. Chaix, C. Moustrou, B. Mandrand and H. Brisset, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 2439–2441.

112 N. Hüsken, G. Gasser, S. D. Köster and N. Metzler-Nolte, *Bioconjugate Chem.*, 2009, **20**, 1578–1586.

113 X. Luo, T. M.-H. Lee and I.-M. Hsing, *Anal. Chem.*, 2008, **80**, 7341–7346.

114 S. C. Hillier, C. G. Frost, A. T. A. Jenkins, H. T. Braven, R. W. Keay, S. E. Flower and J. M. Clarkson, *Bioelectrochemistry*, 2004, **63**, 307–310.

115 S. C. Hillier, S. E. Flower, C. G. Frost, A. T. A. Jenkins, R. Keay, H. Braven and J. Clarkson, *Electrochem. Commun.*, 2004, **6**, 1227–1232.

116 B. J. Marsh, J. Sharp, S. E. Flower, C. G. Frost, Eu. Pat., EP 2655387, 2014.

117 B. J Marsh, C. G. Frost, J. Sharp, Eu. Pat., EP 3055317, 2018.

118 T. J. Kealy and P. L. Pauson, *Nature*, 1951, **168**, 1039–1040.

119 G. Wilkinson, M. Rosenblum, M. C. Whiting and R. B. Woodward, *J. Am. Chem. Soc.*, 1952, **74**, 2125–2126.

120 M. Nakahata, Y. Takashima, H. Yamaguchi and A. Harada, *Nat. Commun.*, 2011, **2**, 511.

121 R. Gómez Arrayás, J. Adrio and J. C. Carretero, *Angew. Chem. Int. Ed.*, 2006, **45**, 7674–7715.

122 R. R. Gagne, C. A. Koval and G. C. Lisensky, *Inorg. Chem.*, 1980, **19**, 2854–2855.

123 G. Oudijk, *Environ. Forensics*, 2010, **11**, 17–49.

124 M. Patra and G. Gasser, *Nat. Rev. Chem.*, 2017, **1**, 66.

125 J.-L. H. a. Duprey and J. H. R. Tucker, *Chem. Lett.*, 2014, **43**, 157–163.

126 S. Goggins, B. J. Marsh, A. T. Lubben and C. G. Frost, *Chem. Sci.*, 2015, **6**, 4978–4985.

127 D. Marquarding, H. Klusacek, G. Gokel, P. Hoffmann and I. Ugi, *J. Am. Chem. Soc.*, 1970, **92**, 5389–5393.

128 L. F. Battelle, R. Bau, G. W. Gokel, R. T. Oyakawa and I. Ugi, *Angew. Chem. Int. Ed.*, 1972, **11**, 138–140.

129 L. F. Battelle, R. Bau, G. W. Gokel, R. T. Oyakawa and I. K. Ugi, *J. Am. Chem. Soc.*, 1973, **95**, 482–486.

130 O. Riant, S. Odile, T. Flessner, S. Taudien and H. B. Kagan, *J. Org. Chem.*, 1997, **62**, 6733–6745.

131 A. N. Nesmeyanov, E. V. Leonova, N. S. Kochetkova and A. I. Makkova, *J. Organomet. Chem.*, 1975, **96**, 275–278.

132 C. Pichon, B. Odell and J. M. Brown, *Chem. Commun.*, 2004, 598–599.

133 R. Sanders and U. T. Mueller-Westerhoff, *J. Organomet. Chem.*, 1996, **512**, 219–224.

134 B. J. Marsh, L. Hampton, S. Goggins and C. G. Frost, *New J. Chem.*, 2014, **38**, 5260–5263.

135 M. J. Robins and J. S. Wilson, *J. Am. Chem. Soc.*, 1981, **103**, 932–933.

136 G. Luoni, C. McGuigan, G. Andrei, R. Snoeck, E. De Clercq and J. Balzarini, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 3791–3796.

137 O. Bidet, C. McGuigan, G. Andrei, R. Snoeck, E. De Clercq and J. Balzarini, *Nucleosides, Nucleotides*, 2003, **22**, 817–819.

138 C. Yu and F. Oberdorfer, *Synlett*, 2000, **1**, 86–88.

139 M. J. Robins and P. J. Barr, *J. Org. Chem.*, 1983, **48**, 1854–1862.

140 S. Meneni, I. Ott, C. D. Sergeant, A. Sniady, R. Gust and R. Dembinski, *Bioorg. Med. Chem.*, 2007, **15**, 3082–3088.

141 W. A. Cristofoli, L. I. Wiebe, E. De Clercq, G. Andrei, R. Snoeck, J. Balzarini and E. E. Knaus, *J. Med. Chem.*, 2007, **50**, 2851–2857.

142 G. T. Crisp and B. L. Flynn, *J. Org. Chem.*, 1993, **58**, 6614–6619.

143 D. Graham, J. A. Parkinson and T. Brown, *J. Chem. Soc. Perkin Trans. 1*, 1998, 1131–1138.

144 D. J. Hurley and Y. Tor, *J. Am. Chem. Soc.*, 2002, **124**, 3749–3762.

145 D. J. Hurley and Y. Tor, *J. Am. Chem. Soc.*, 1998, **120**, 2194–2195.

146 M. Münzel, D. Globisch, C. Trindler and T. Carell, *Org. Lett.*, 2010, **12**, 5671–5673.

147 P. K. Chang and A. D. Welch, *J. Med. Chem.*, 1963, **6**, 428–430.

148 J. Dadová, M. Vrábel, M. Adámik, M. Brázdová, R. Pohl, M. Fojta and M. Hocek, *Chem. Eur. J.*, 2015, **21**, 16091–16102.

149 M. Münzel, C. Szeibert, A. F. Glas, D. Globisch and T. Carell, *J. Am. Chem. Soc.*, 2011, **133**, 5186–5189.

150 M. S. Inkpen, S. Du, M. Driver, T. Albrecht and N. J. Long, *Dalton Trans.*, 2013, **42**, 2813–6.

151 G. V. M. Sharma, T. Rajendra Prasad and A. K. Mahalingam, *Tetrahedron Lett.*, 2001, **42**, 759–761.

152 V. M. Vogt, *Eur. J. Biochem.*, 1973, **33**, 192–200.

153 ThermoFisher, User Guide: S1 Nuclease, https://www.thermofisher.com/order/catalog/product/EN0321, (accessed 4 April 2018).

154 C. Hatt, M. E. Ward and I. N. Clarke, *Nucleic Acids Res.*, 1988, **16**, 4053–4067.

155 B. J. Marsh, C. G. Frost, D. Pearce, Eu. Pat., EP 2864342 (Pending), 2012.

156 S. Sato, M. Tsueda and S. Takenaka, *J. Organomet. Chem.*, 2010, **695**, 1858–1862.

157 H. Sun, W. Chen and A. E. Kaifer, *Organometallics*, 2006, **25**, 1828–1830.

158 J. Alvarez and A. E. Kaifer, *Organometallics*, 1999, **18**, 5733–5734.

159 J. Alvarez, T. Ren and A. E. Kaifer, *Organometallics*, 2001, **20**, 3543–3549.

160 N. C. Tice, S. Parkin and J. P. Selegue, *J. Organomet. Chem.*, 2007, **692**, 791–800.

161 J. L. Kerr, J. S. Landells, D. S. Larsen, B. H. Robinson and J. Simpson, *J. Chem. Soc. Dalt. Trans.*, 2000, 1411–1417.

162 M. S. Inkpen, S. Du, M. Hildebrand, A. J. P. White, N. M. Harrison, T. Albrecht and N. J. Long, *Organometallics*, 2015, **34**, 5461–5469.

163 CDC Fact Sheet: Incidence , Prevalence , and Cost of Sexually Transmitted Infections in the United States, https://www.cdc.gov/std/stats/sti-estimates-fact-sheet-feb-2013.pdf, (accessed 13 April 2018).

164 L. Newman, J. Rowley, S. Vander Hoorn, N. S. Wijesooriya, M. Unemo, N. Low, G. Stevens, S. Gottlieb, J. Kiarie and M. Temmerman, *PLoS One*, 2015, **10**, e0143304.

165 M. Unemo, C. S. Bradshaw, J. S. Hocking, H. J. C. de Vries, S. C. Francis, D. Mabey, J. M. Marrazzo, G. J. B. Sonder, J. R. Schwebke, E. Hoornenborg, R. W. Peeling, S. S. Philip, N. Low and C. K. Fairley, *Lancet Infect. Dis.*, 2017, **17**, e235–e279.

166 L. Greer and G. D. Wendel, Jr, *Infect. Dis. Clin. North Am.*, 2008, **22**, 601–617.

167 S. R. Galvin and M. S. Cohen, *Nat. Rev. Microbiol.*, 2004, **2**, 33–42.

168 Public Health England, Table 1: STI diagnoses & rates in England by gender, 2007 - 2016, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/626359/2016_Table_1_STI_diagnoses___rates_in_England_by_gender.pdf, (accessed 10 May 2018).

169 Public Health England, Table 8 : Attendances by gender, sexual risk & age group, 2012 - 2016, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/617052/2016_Table_8_Attendances_by_gender__sexual_risk_and_age_group.pdf, (accessed 10 May 2018).

170 Y. P. Yin, R. W. Peeling, X. S. Chen, K. L. Gong, H. Zhou, W. M. Gu, H. P. Zheng, Z. S. Wang, G. Yong, W. L. Cao, M. Q. Shi, W. H. Wei, X. Q. Dai, X. Gao, Q. Chen and D. Mabey, *Sex. Transm. Infect.*, 2006, **82**, 33–37.

171 A. S. Benzaken, E. G. Galban, W. Antunes, J. C. Dutra, R. W. Peeling, D. Mabey and A. Salama, *Sex. Transm. Infect.*, 2006, **82**, 26–28.

172 W. Huang, C. A. Gaydos, M. R. Barnes, M. Jett-Goheen and D. R. Blake, *Sex. Transm. Infect.*, 2013, **89**, 108–14.

173 T. U. Berendonk, C. M. Manaia, C. Merlin, D. Fatta-Kassinos, E. Cytryn, F. Walsh, H. Bürgmann, H. Sørum, M. Norström, M. N. Pons, N. Kreuzinger, P. Huovinen, S. Stefani, T. Schwartz, V. Kisand, F. Baquero and J. L. Martinez, *Nat. Rev. Microbiol.*, 2015, **13**, 310–317.

174 G. Brook, *Sex. Transm. Infect.*, 2015, **91**, 539–544.

175 K. M. E. Turner, J. Round, P. Horner, J. Macleod, S. Goldenberg, A. Deol and E. J. Adams, *Sex. Transm. Infect.*, 2014, **90**, 104–111.

176 H. Kelly, C. E. M. Coltart, N. Pant Pai, J. D. Klausner, M. Unemo, I. Toskin and R. W. Peeling, *Sex. Transm. Infect.*, 2017, **93**, S22–S30.

177 L. A. Watchirs Smith, R. Hillman, J. Ward, D. M. Whiley, L. Causer, S. Skov, B. Donovan, J. Kaldor and R. Guy, *Sex. Transm. Infect.*, 2013, **89**, 320–326.

178 A. St John and C. P. Price, *Clin. Biochem. Rev.*, 2014, **35**, 155–167.

179 C. A. Gaydos and J. Hardick, *Expert Rev. Anti Infect. Ther.*, 2014, **12**, 657–672.

180 P. Vickerman, C. Watts, M. Alary, D. Mabey and R. W. Peeling, *Sex. Transm. Infect.*, 2003, **79**, 363–367.

181 H. Braven, R. Keay, Eu. Pat., EP 1481083, 2006.

182 D. Filer, C. Ferrao, S. Chadwick, Eu. Pat., EP 3194617 (Pending), 2014.

183 D. Alderstein, D. M. Pearce, Eu. Pat., EP 3022318, 2013.

184 B. Arlett, J. K. Taylor, K. D. Neale, T. C. E. Mullarkey, T. R. K. Edwards, US Pat., 9,816,135-B2, 2017.

185 D. M. Pearce, D. P. Shenton, J. Holden and N. W. Street, *IEEE Trans. Biomed. Eng.*, 2011, **58**, 755–758.

186 K. Shinozaki and T. Okazaki, *Nucleic Acids Res.*, 1978, **5**, 4245–4261.

187 L. A. Lewis, A. F. Gillaspy, R. E. McLaughlin, M. Gipson, T. F. Ducey, T. Ownbey, K. Hartman, C. Nydick, M. B. Carson, J. Vaughn, C. Thomson, L. Song, S. Lin, X. Yuan, F. Najar, M. Zhan, Q. Ren, H. Zhu, S. Qi, S. M. Kenton, H. Lai, J. D. White, S. Clifton, B. A. Roe and D. W. Dyer, *GenBank*, 2003, AE004969.

188 P. G. Mitsis and J. G. Kwagh, *Nucleic Acids Res.*, 1999, **27**, 3057–3063.

189 J. W. Little, *J. Biol. Chem.*, 1977, **242**, 679–686.

190 M. Storm, *Unpublished Work*.

191 K. S. Sriprakash, N. Lundh, M. M.-O. Huh and C. M. Radding, *J. Biol. Chem.*, 1975, **250**, 5438–5455.

192 H. E. Gottlieb, V. Kotlyar and A. Nudelman, *J. Org. Chem.*, 1997, **62**, 7512–7515.

193 D. J. Hurley and Y. Tor, *J. Am. Chem. Soc.*, 1998, **120**, 2194–2195.

194 N. K. Andersen, H. Døssing, F. Jensen, B. Vester and P. Nielsen, *J. Org. Chem*, 2011, **76**, 6177–6187.

195 R. A. I. Abou-Elkhair and T. L. Netzel, *Nucleosides, Nucleotides*, 2005, **24**, 85–110.

196 Y. Zhang, X. Yue, B. Kim, S. Yao and K. D. Belfield, *Chem. Eur. J.*, 2014, **20**, 7249–7253.

197 T.-Y. Dong and L.-L. Lai, *J. Organomet. Chem.*, 1996, **509**, 131–134.

198 P. Argyropoulos, F. Bergeret, C. Pardin, J. M. Reimer, A. Pinto, C. N. Boddy and T. M. Schmeing, *Biochim. Biophys. Acta*, 2016, **1860**, 486–497.

199 F. Coutrot and E. Busseron, *Chem. Eur. J.*, 2009, **15**, 5186–5190.

200 A. Barnard, K. Long, D. J. Yeo, J. A. Miles, V. Azzarito, G. M. Burslem, P. Prabhakaran, T. A. Edwards and A. J. Wilson, *Org. Biomol. Chem.*, 2014, **12**, 6794–6799.

201 J. Federič and Š. Toma, *Collect. Czechoslov. Chem. Commun.*, 1987, **52**, 174–181.

202 F. Geisler and G. Helmchen, *Synthesis*, 2006, **2006**, 2201–2205.

Appendix 1

Crystal data and structure refinement for **70**

Identification code	e15cfg1	
Empirical formula	C28 H35 Fe2 N O	
Formula weight	513.27	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	P 21/n	
Unit cell dimensions	a = 11.2994(3) Å	α= 90°.
	b = 16.8718(5) Å	β= 98.365(3)°.
	c = 12.6013(4) Å	γ= 90°.
Volume	2376.77(12) Å3	
Z	4	
Density (calculated)	1.434 Mg/m^3	
Absorption coefficient	1.241 mm^{-1}	
F(000)	1080	
Crystal size	0.500 x 0.350 x 0.060 mm^3	
Theta range for data collection	3.484 to 29.339°.	
Index ranges	-15<=h<=12, -21<=k<=17, -14<=l<=16	
Reflections collected	18871	
Independent reflections	5612 [R(int) = 0.0301]	
Completeness to theta = 25.242°	99.7 %	
Absorption correction	Semi-empirical from equivalents	
Max. and min. transmission	1.00000 and 0.85628	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	5612 / 0 / 293	
Goodness-of-fit on F^2	1.031	
Final R indices [I>2sigma(I)]	R1 = 0.0316, wR2 = 0.0623	
R indices (all data)	R1 = 0.0458, wR2 = 0.0668	
Extinction coefficient	n/a	
Largest diff. peak and hole	0.348 and -0.318 $e.Å^{-3}$	

Appendix 2

Crystal data and structure refinement for **73**

Identification code	k15cgf5	
Empirical formula	C28 H33 Br2 Fe2 N O	
Formula weight	671.07	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	P21/n	
Unit cell dimensions	a = 12.41410(10) Å	α= 90°.
	b = 16.8699(2) Å	β= 99.5089(5)°.
	c = 12.80360(10) Å	γ = 90°.
Volume	2644.55(4) Å3	
Z	4	
Density (calculated)	1.685 Mg/m^3	
Absorption coefficient	4.141 mm^{-1}	
F(000)	1352	
Crystal size	0.500 x 0.300 x 0.200 mm3	
Theta range for data collection	3.212 to 30.019°.	
Index ranges	-17<=h<=17, -23<=k<=23, -18<=l<=18	
Reflections collected	55114	
Independent reflections	7696 [R(int) = 0.0621]	
Completeness to theta = 25.242°	99.6 %	
Absorption correction	Semi-empirical from equivalents	
Max. and min. transmission	0.455 and 0.229	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	7696 / 1 / 311	
Goodness-of-fit on F^2	1.029	
Final R indices [I>2sigma(I)]	R1 = 0.0345, wR2 = 0.0836	
R indices (all data)	R1 = 0.0465, wR2 = 0.0900	
Extinction coefficient	n/a	
Largest diff. peak and hole	0.773 and -1.059 $e.Å^{-3}$	

Appendix 3

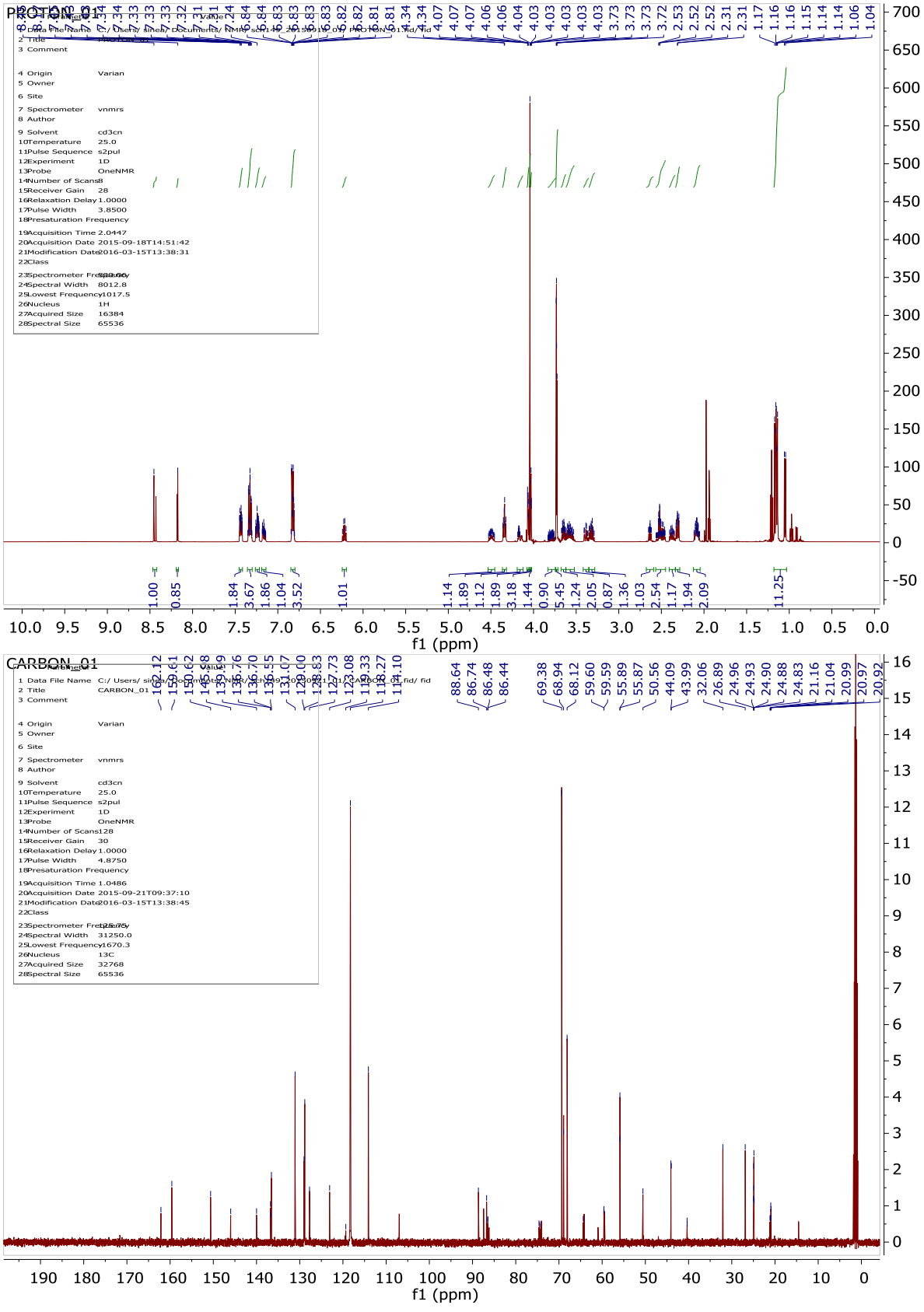
Crystal data and structure refinement for **72**

Identification code	k15cgf3	
Empirical formula	C28 H35 N O Ru2	
Formula weight	603.71	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Triclinic	
Space group	P-1	
Unit cell dimensions	a = 10.31700(10) Å	α= 105.4920(6)°.
	b = 11.00000(10) Å	β= 98.2650(6)°.
	c = 11.0550(2) Å	γ = 99.1990(6)°.
Volume	1170.32(3) $Å^3$	
Z	2	
Density (calculated)	1.713 Mg/m^3	
Absorption coefficient	1.312 mm^{-1}	
F(000)	612	
Crystal size	0.350 x 0.250 x 0.200 mm^3	
Theta range for data collection	3.073 to 30.029°.	
Index ranges	-14<=h<=14, -15<=k<=14, -15<=l<=15	
Reflections collected	26807	
Independent reflections	6812 [R(int) = 0.0494]	
Completeness to theta = 25.242°	99.8 %	
Absorption correction	Semi-empirical from equivalents	
Max. and min. transmission	0.773 and 0.687	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	6812 / 0 / 293	
Goodness-of-fit on F^2	1.106	
Final R indices [I>2sigma(I)]	R1 = 0.0325, wR2 = 0.0686	
R indices (all data)	R1 = 0.0442, wR2 = 0.0738	
Extinction coefficient	n/a	
Largest diff. peak and hole	1.516 and -1.467 $e.Å^{-3}$	

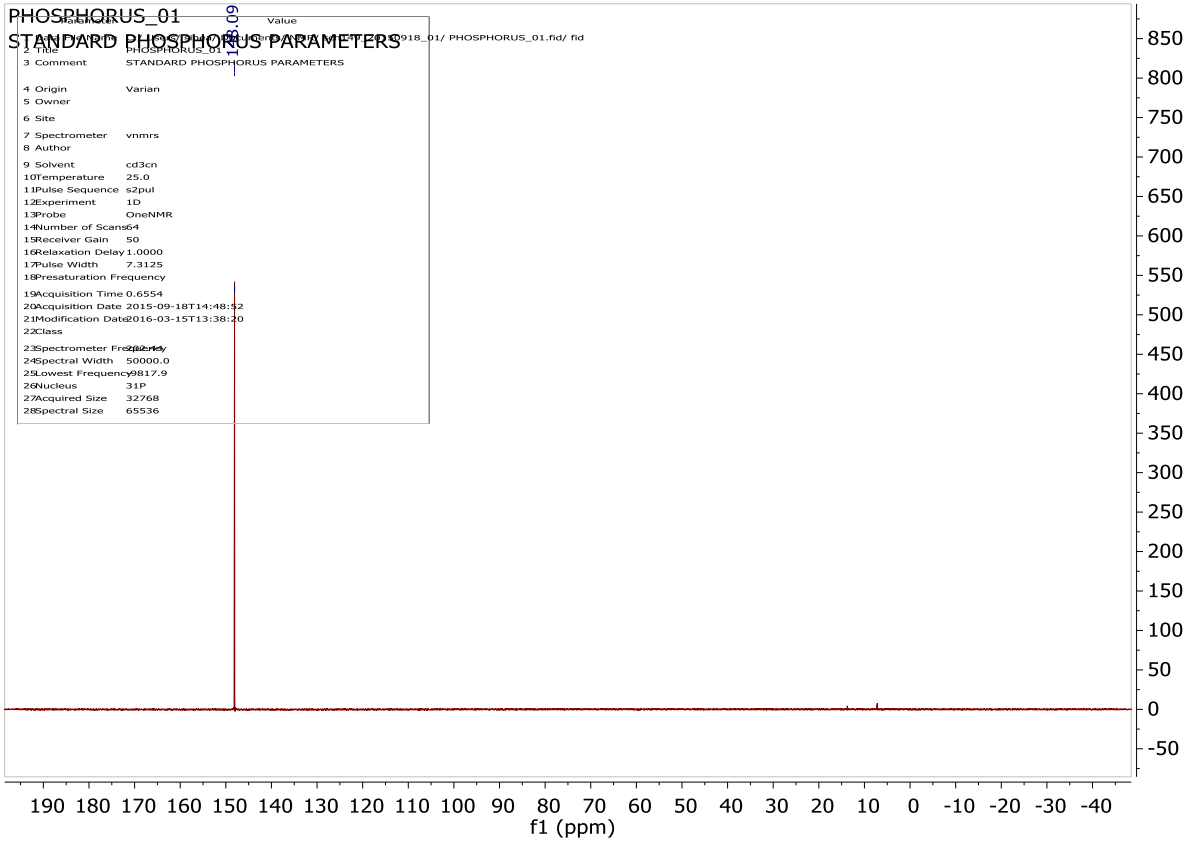
Appendix 4

NMR spectra for phosphoramidites incorporated into oligonucleotides

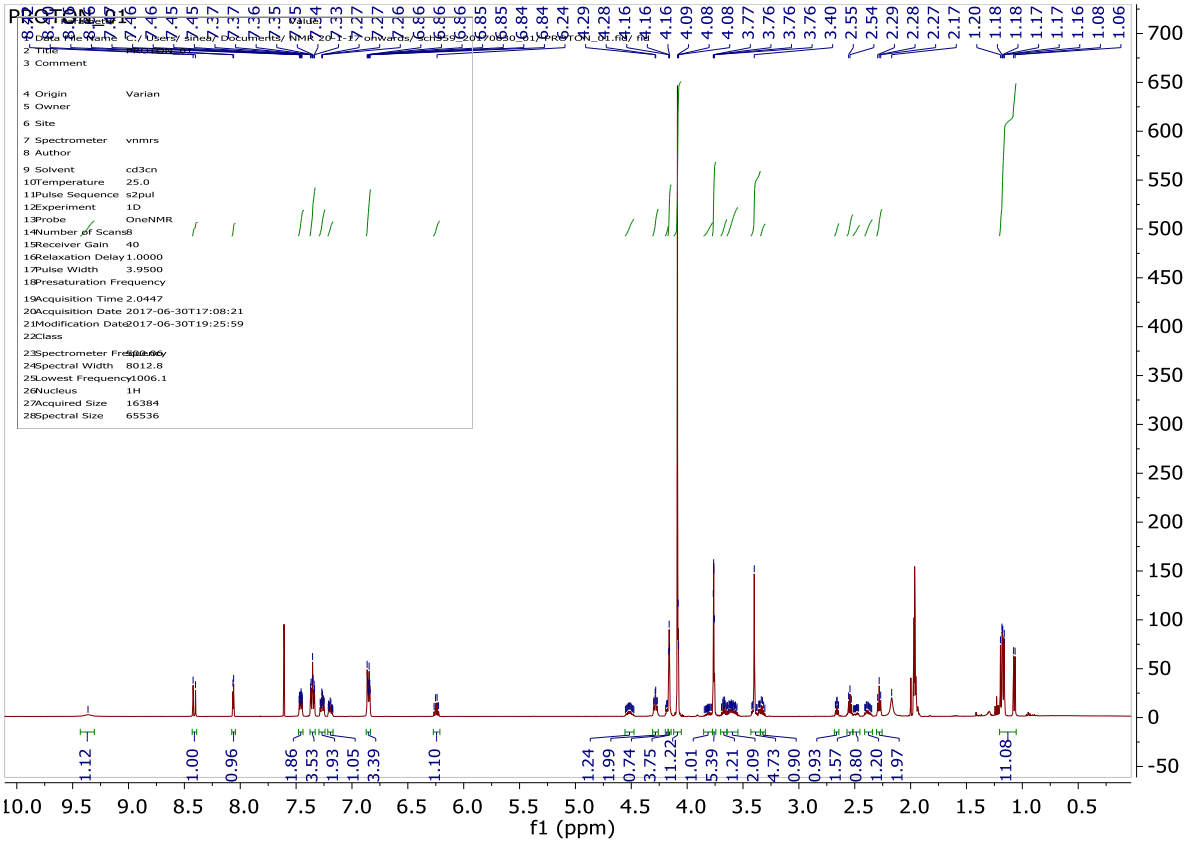
68

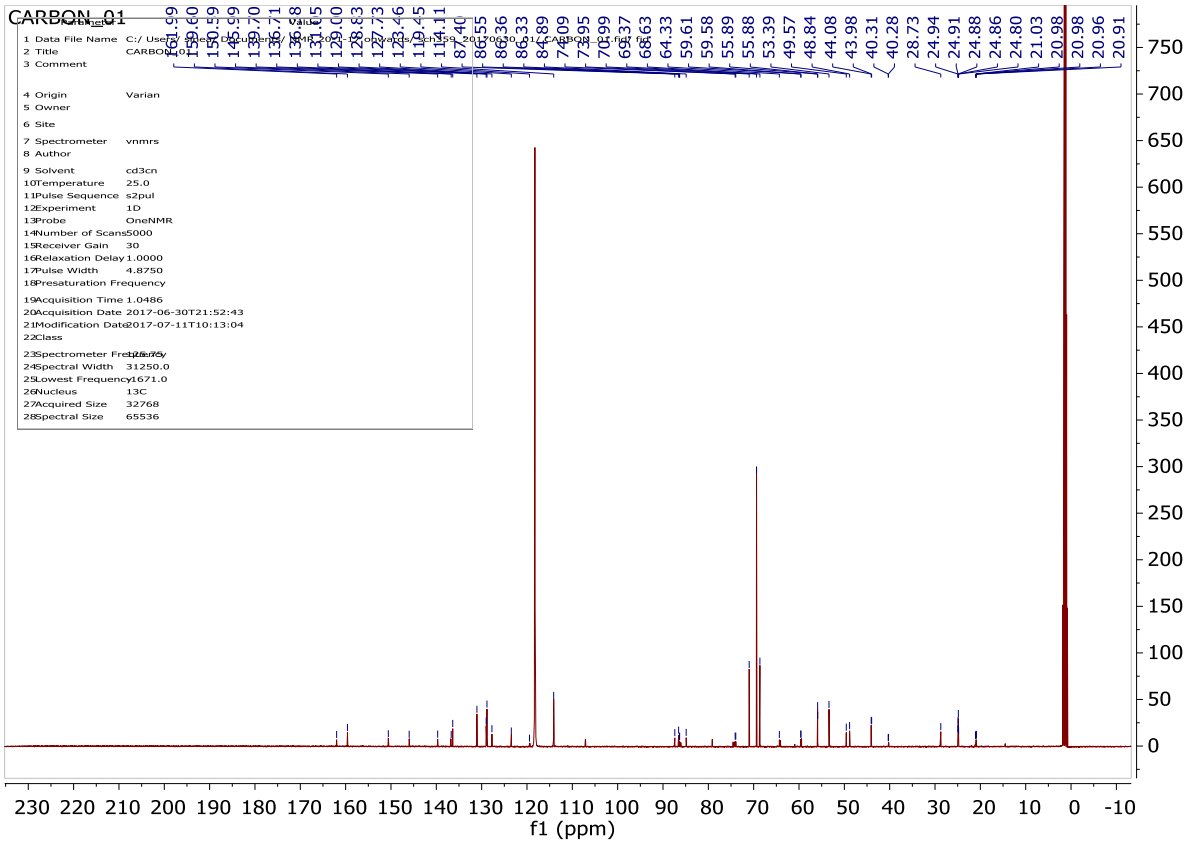


PHOSPHORUS_01
STANDARD PHOSPHORUS PARAMETERS

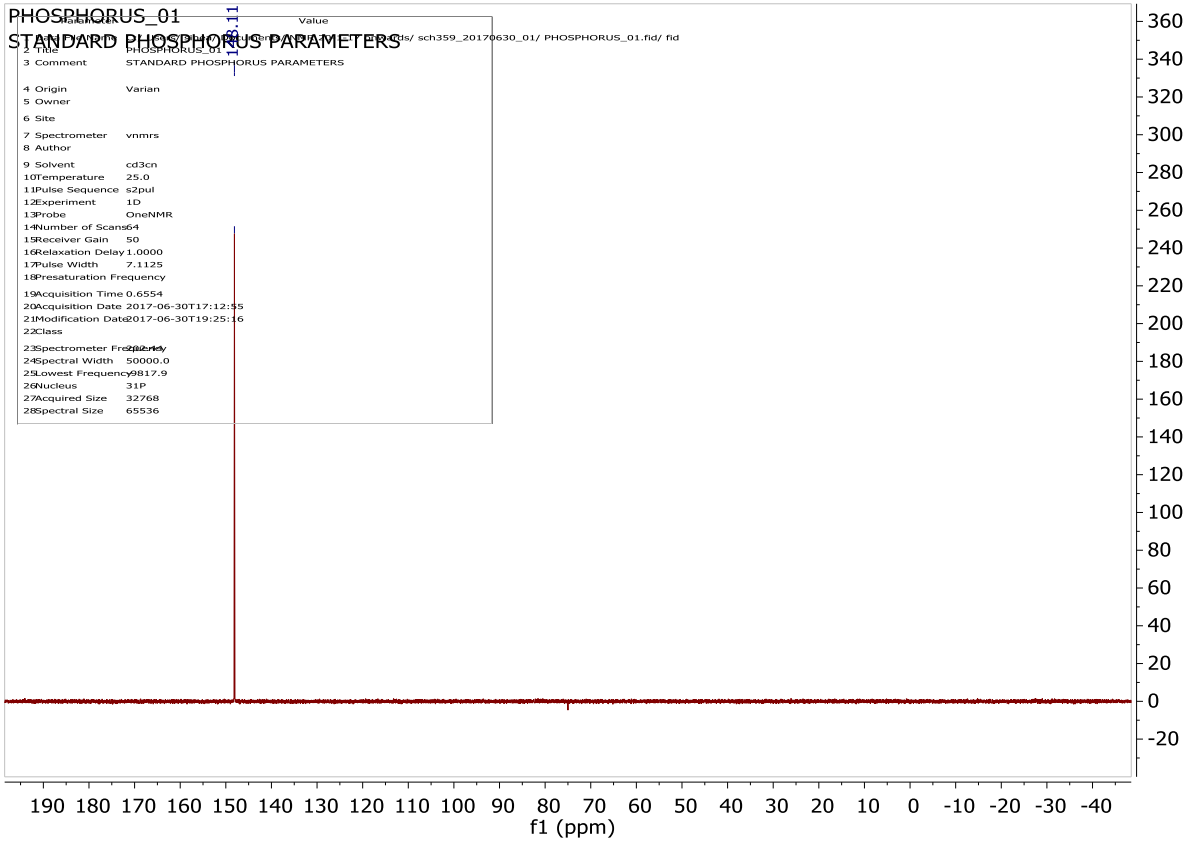


108

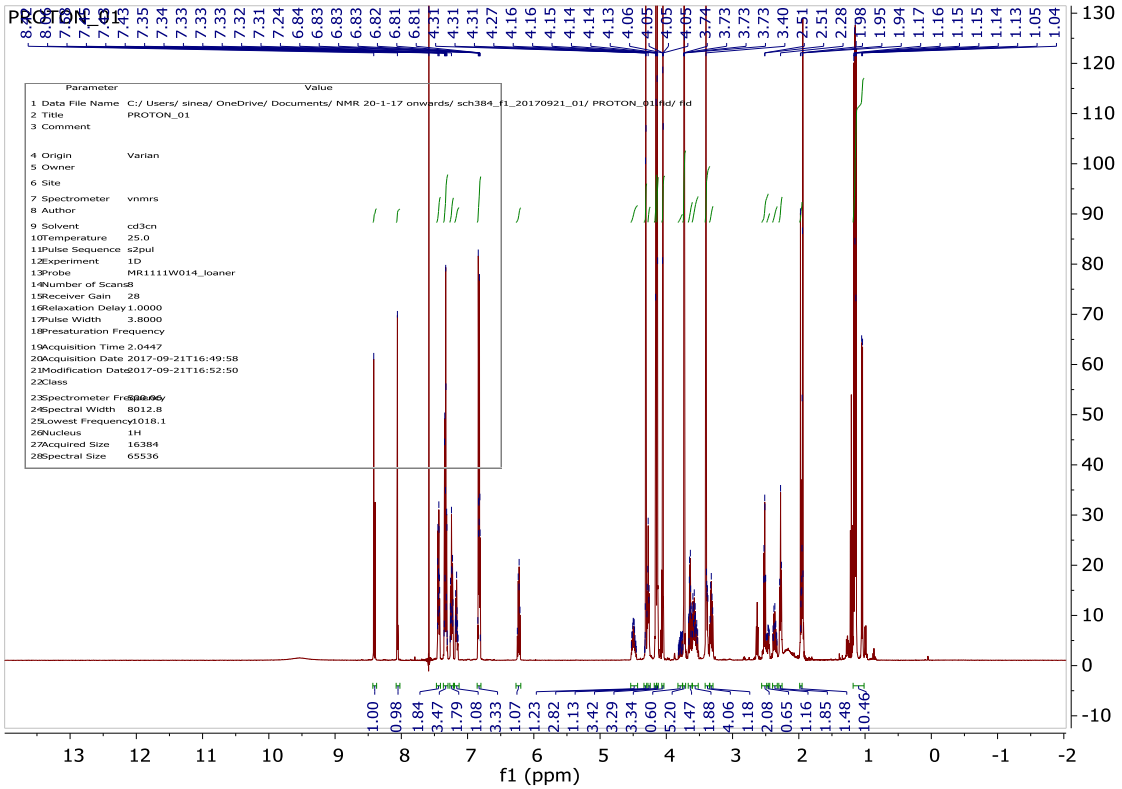


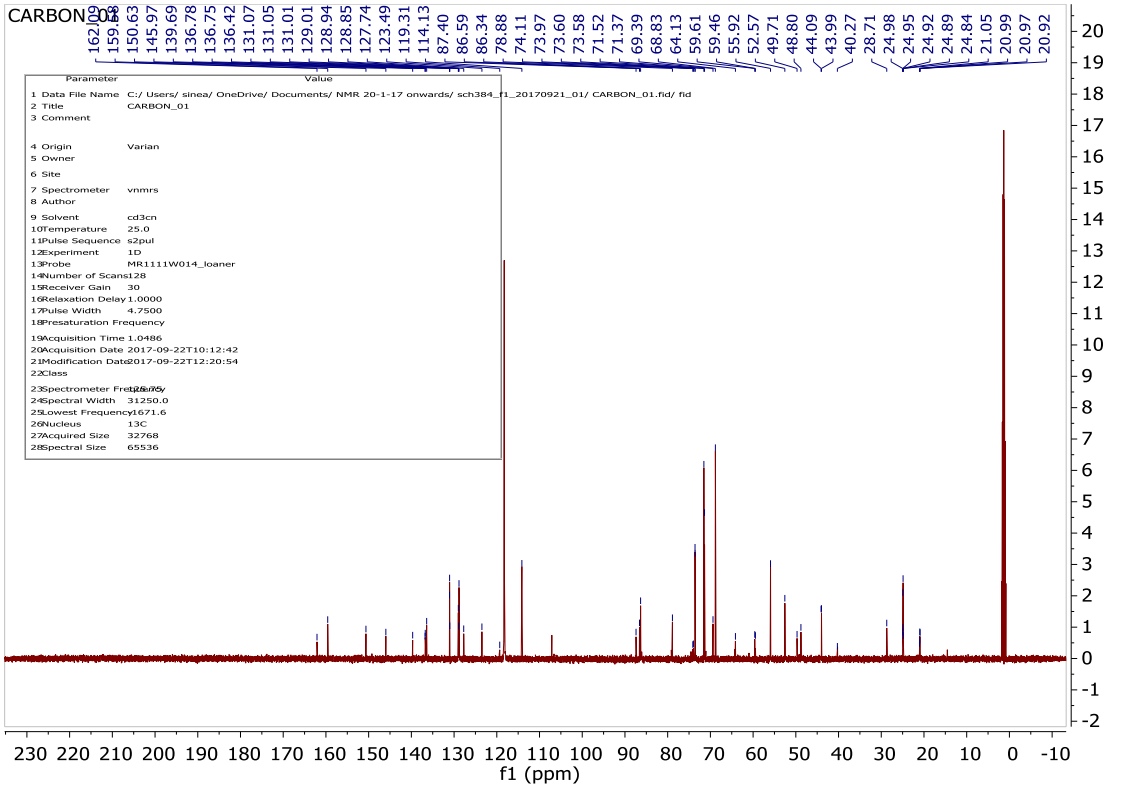


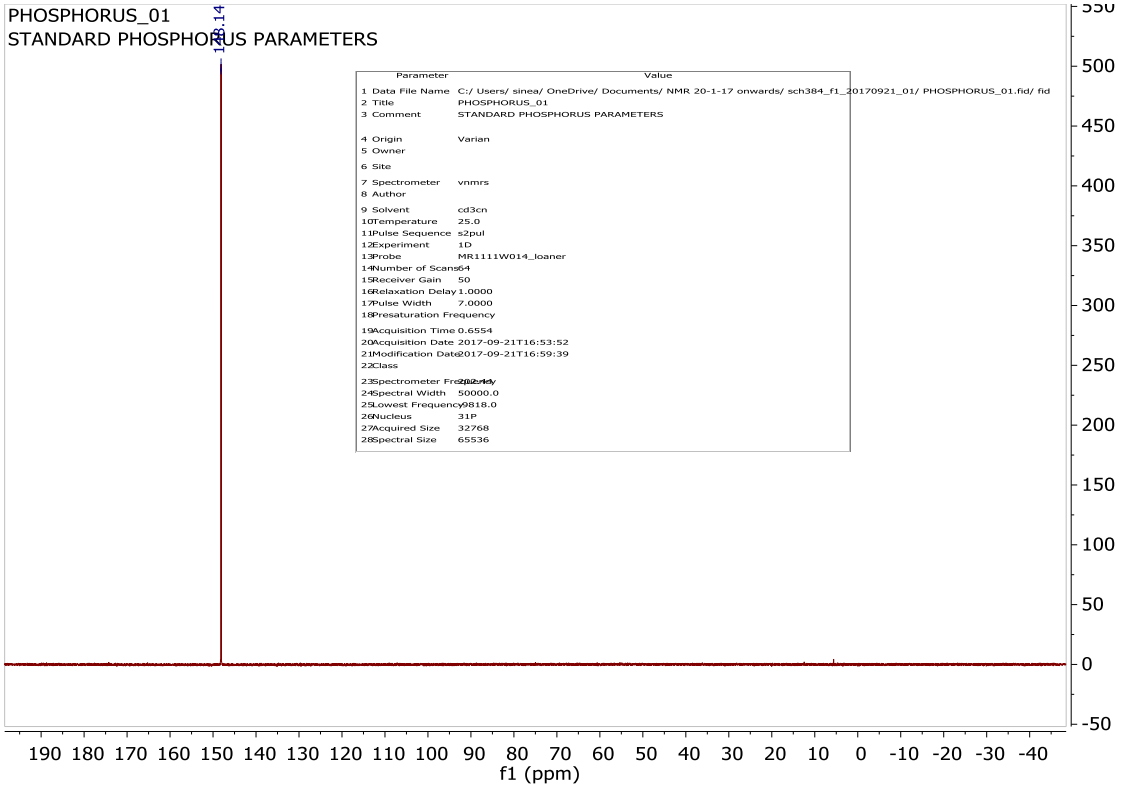
PHOSPHORUS_01
STANDARD PHOSPHORUS PARAMETERS

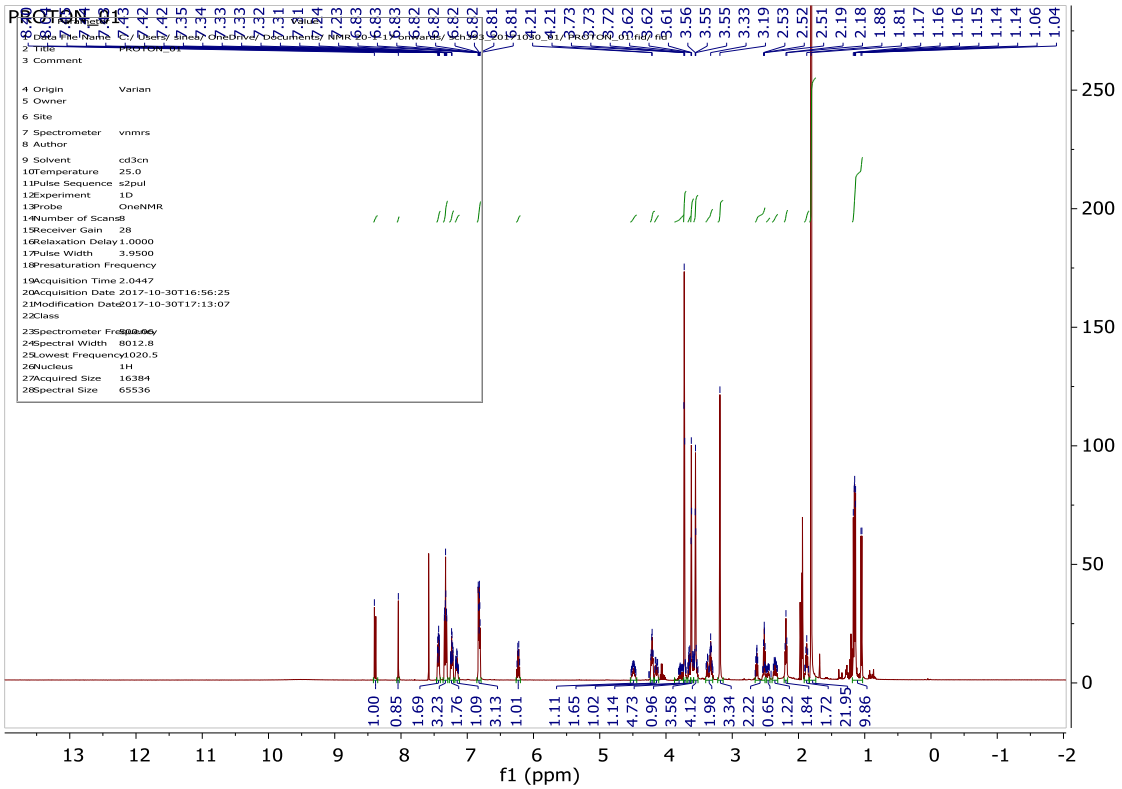


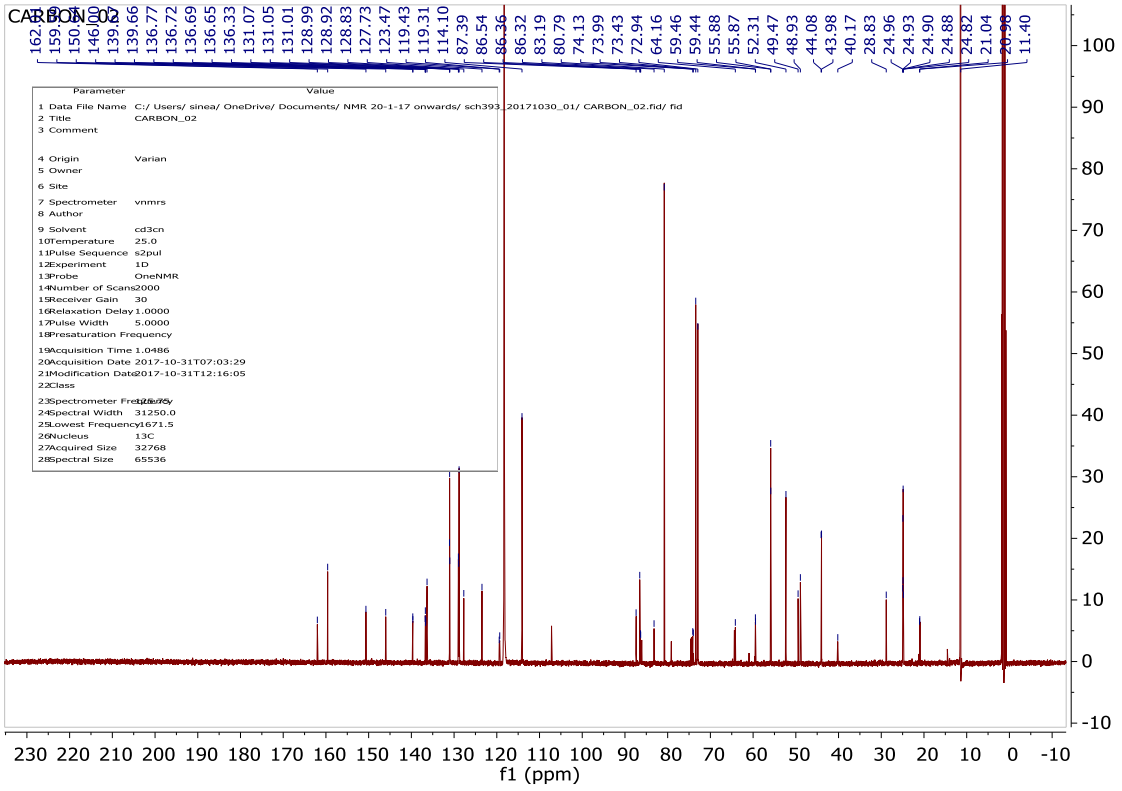
110

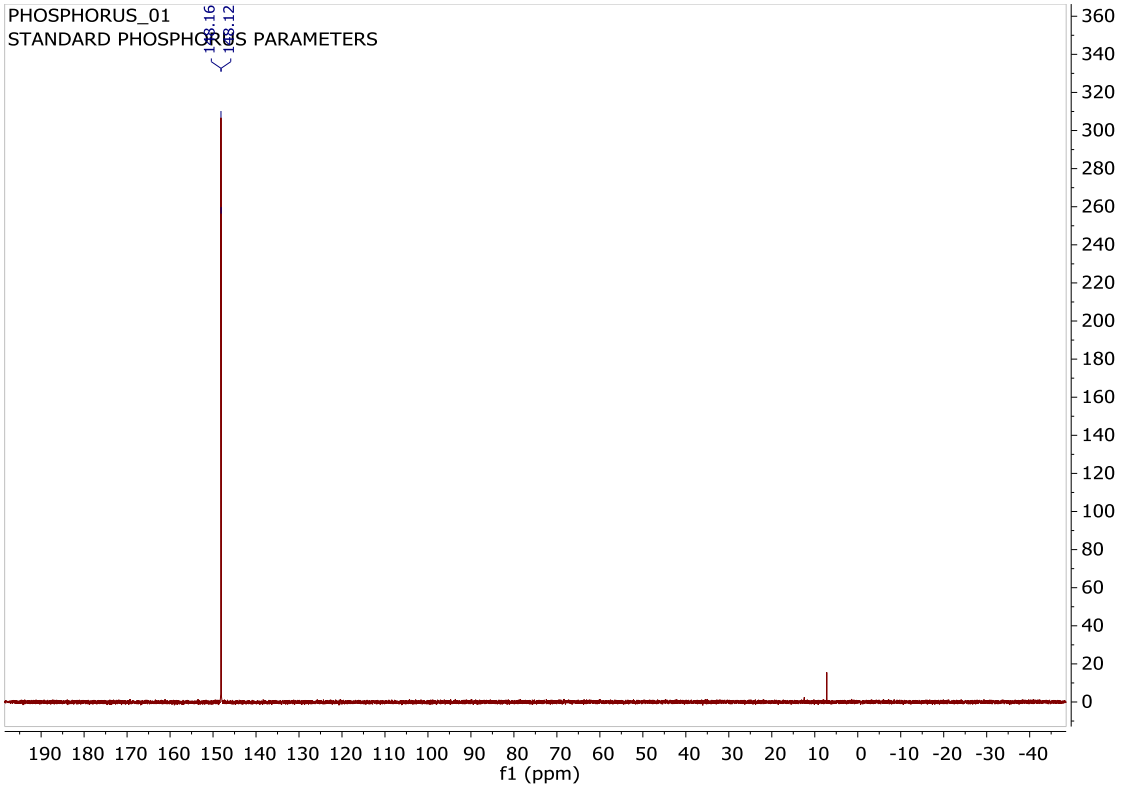





PHOSPHORUS_01
STANDARD PHOSPHORUS PARAMETERS
128.14
Parameter Value
1 Data File Name C:/ Users/ sinea/ OneDrive/ Documents/ NMR 20-1-17 onwards/ sch384_f1_20170921_01/ PHOSPHORUS_01.fid/ fid
2 Title PHOSPHORUS_01
3 Comment STANDARD PHOSPHORUS PARAMETERS
4 Origin Varian
5 Owner
6 Site
7 Spectrometer vnmrs
8 Author
9 Solvent cd3cn
10 Temperature 25.0
11 Pulse Sequence s2pul
12 Experiment 1D
13 Probe MR1111W014_loaner
14 Number of Scans 64
15 Receiver Gain 50
16 Relaxation Delay 1.0000
17 Pulse Width 7.0000
18 Presaturation Frequency
19 Acquisition Time 0.6554
20 Acquisition Date 2017-09-21T16:53:52
21 Modification Date 2017-09-21T16:59:39
22 Class
23 Spectrometer Freq
24 Spectral Width 50000.0
25 Lowest Frequency -9818.0
26 Nucleus 31P
27 Acquired Size 32768
28 Spectral Size 65536
550
500
450
400
350
300
250
200
150
100
50
0
-50
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10 -20 -30 -40
f1 (ppm)


PROTON_01
f1 (ppm)

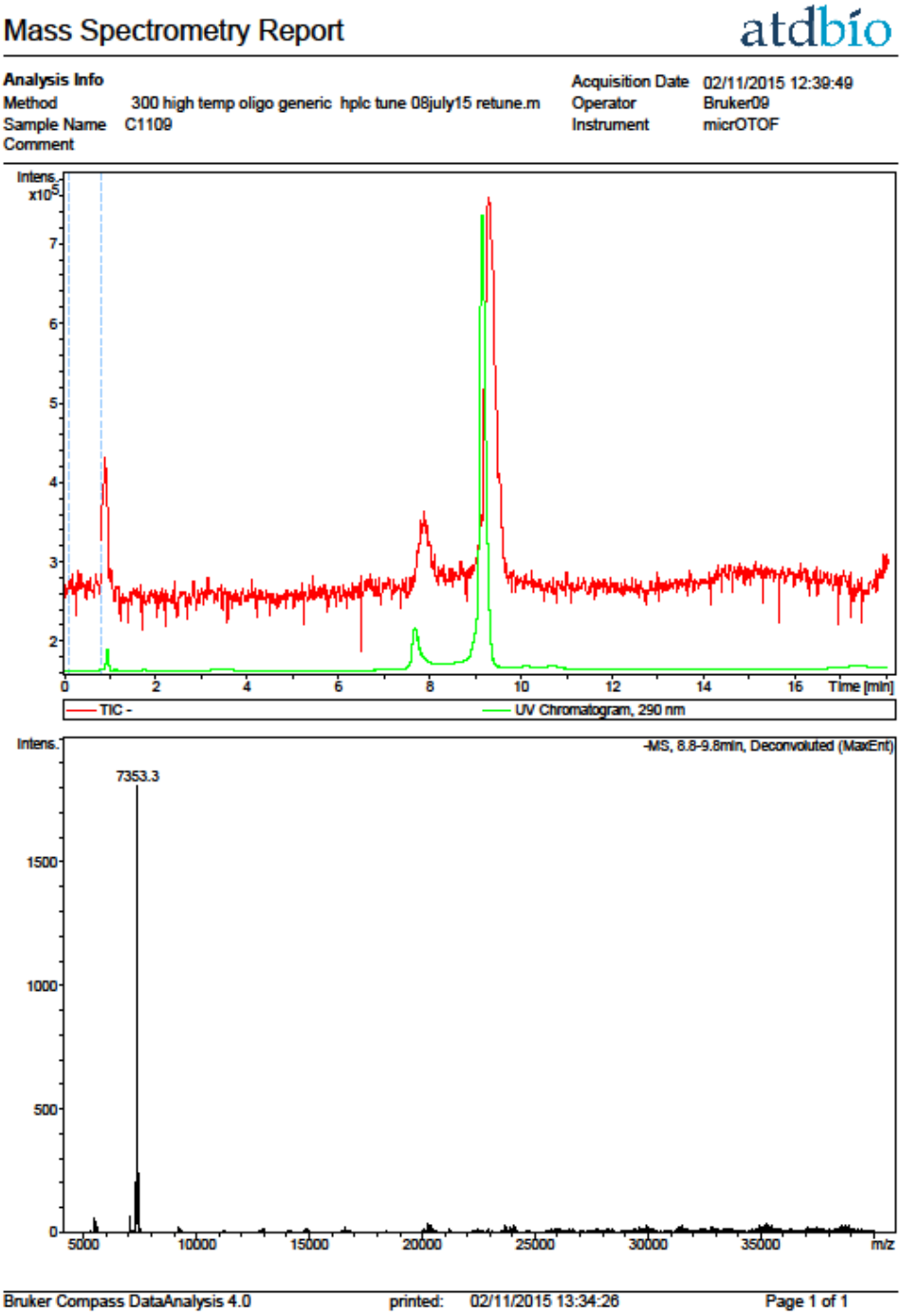

CARBON_02
f1 (ppm)


PHOSPHORUS_01
STANDARD PHOSPHORUS PARAMETERS
150.16
150.12
360
340
320
300
280
260
240
220
200
180
160
140
120
100
80
60
40
20
0
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10 -20 -30 -40
f1 (ppm)

Appendix 5

HPLC-MS Spectra for oligonucleotide probes

IC-C



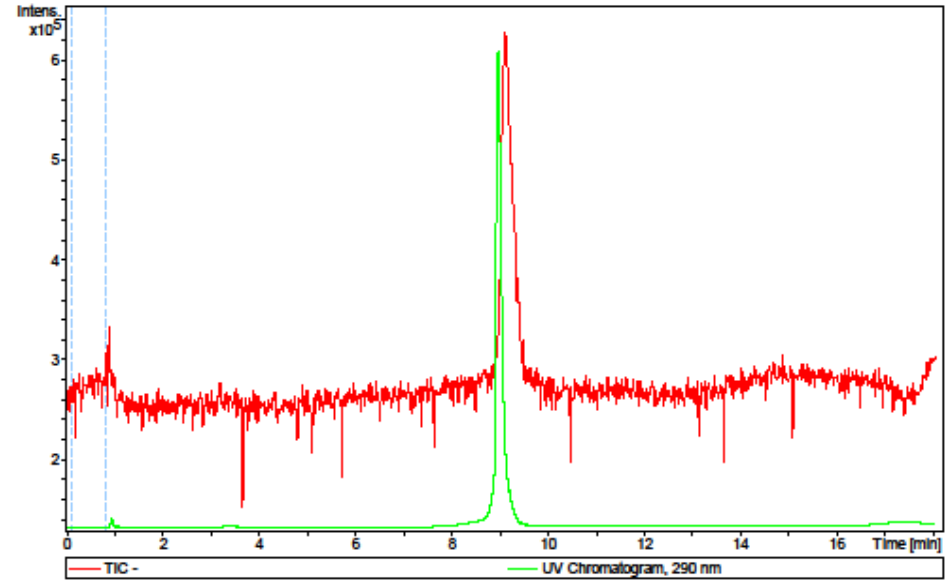
IC-1

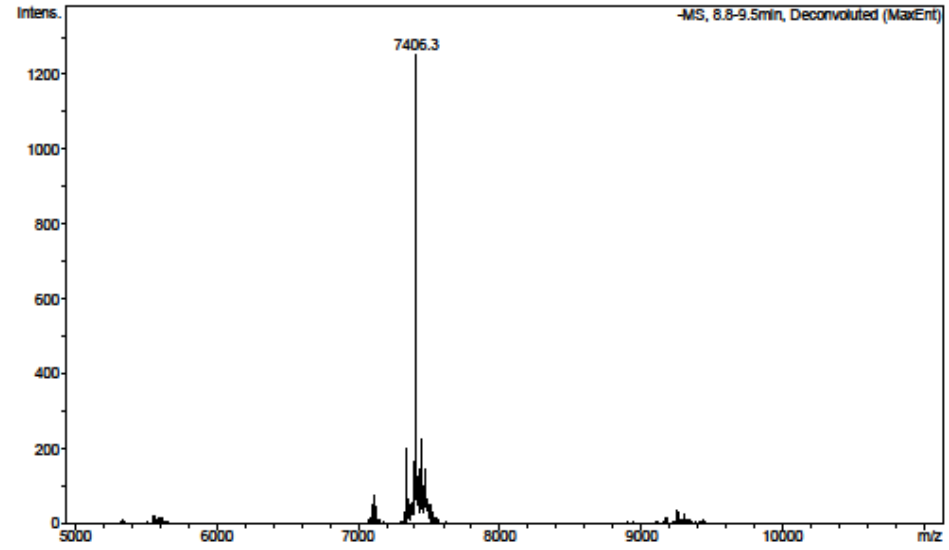
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp oligo generic hplc tune 08july15 retune.m	Acquisition Date	02/11/2015 13:55:36
Sample Name	C1110	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 02/11/2015 16:18:51 Page 1 of 1

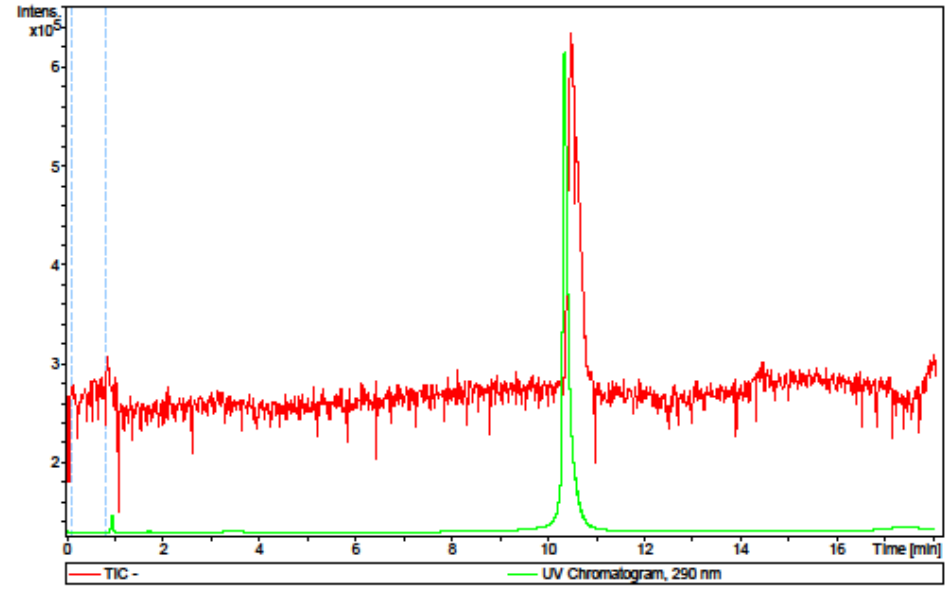
IC-2

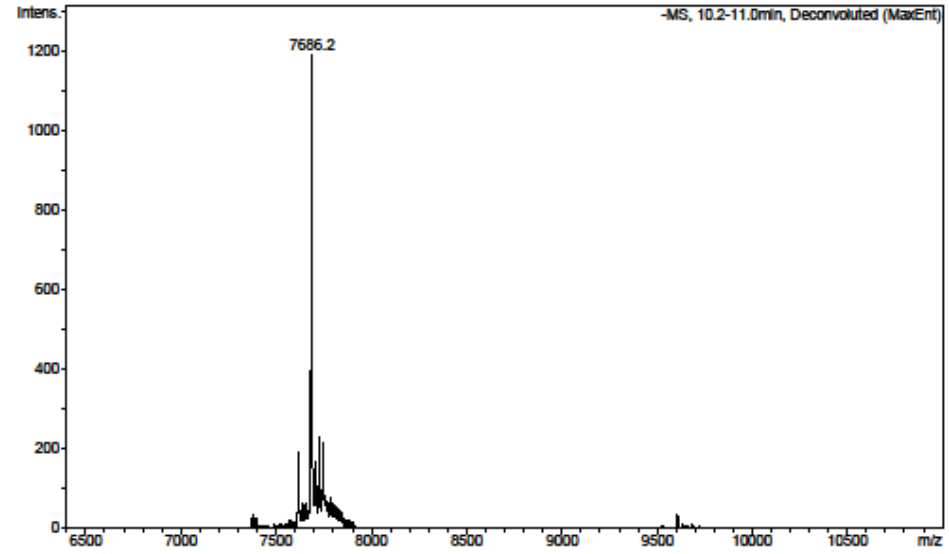
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp oligo generic hplc tune 08july15 retune.m	Acquisition Date	02/11/2015 14:14:32
Sample Name	C1111	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 02/11/2015 16:20:46 Page 1 of 1

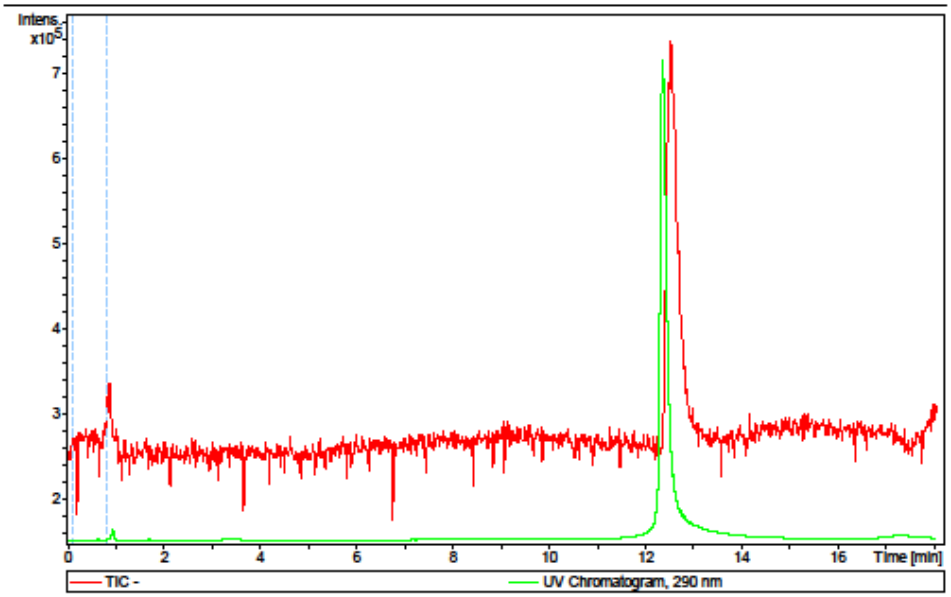
IC-3

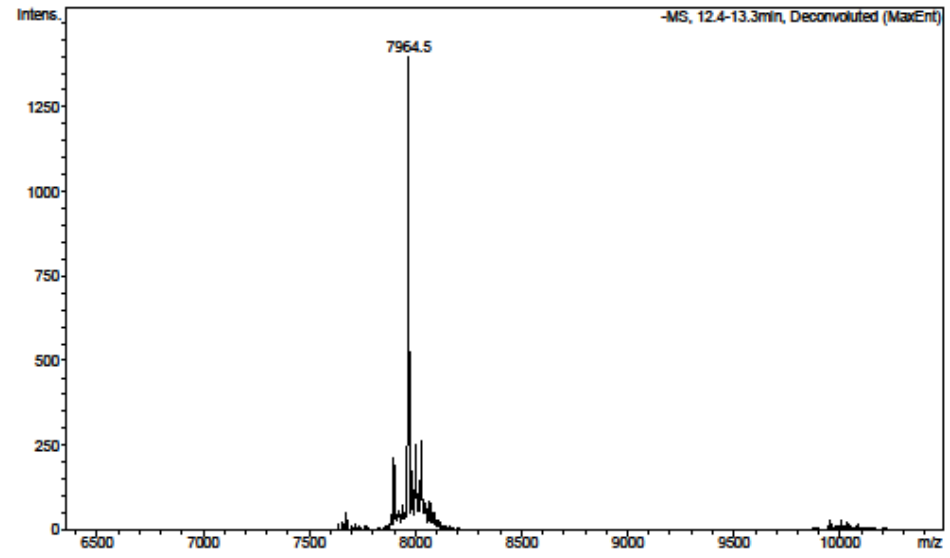
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp oligo generic hplc tune 08july15 retune.m	Acquisition Date	02/11/2015 14:52:30
Sample Name	C1112	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 02/11/2015 16:22:50 Page 1 of 1

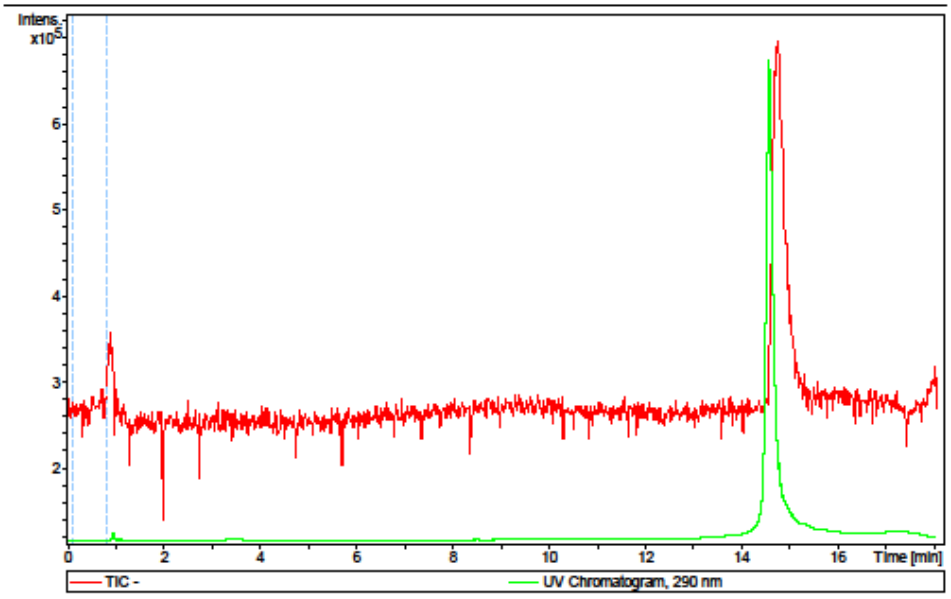
IC-4

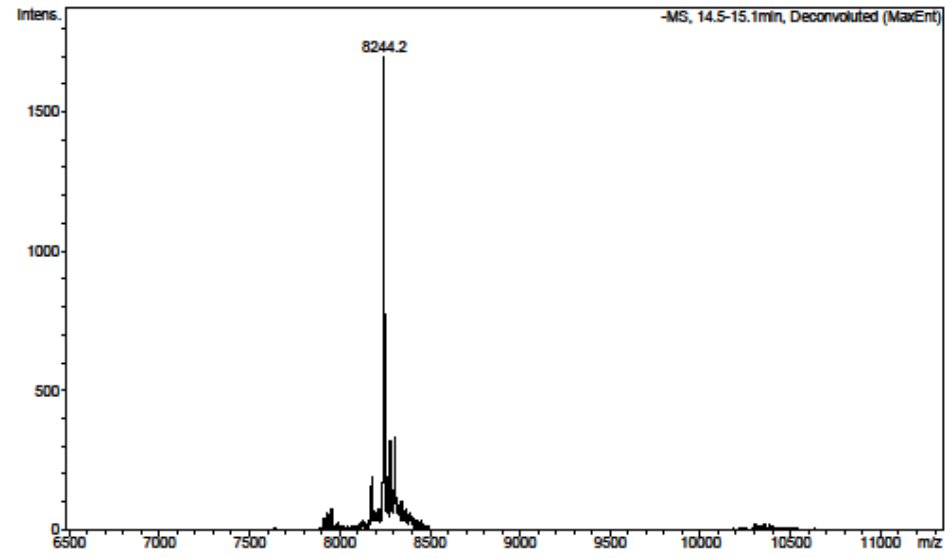
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp oligo generic hplc tune 08july15 retune.m	Acquisition Date	02/11/2015 15:11:23
Sample Name	C1113	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 02/11/2015 16:24:30 Page 1 of 1

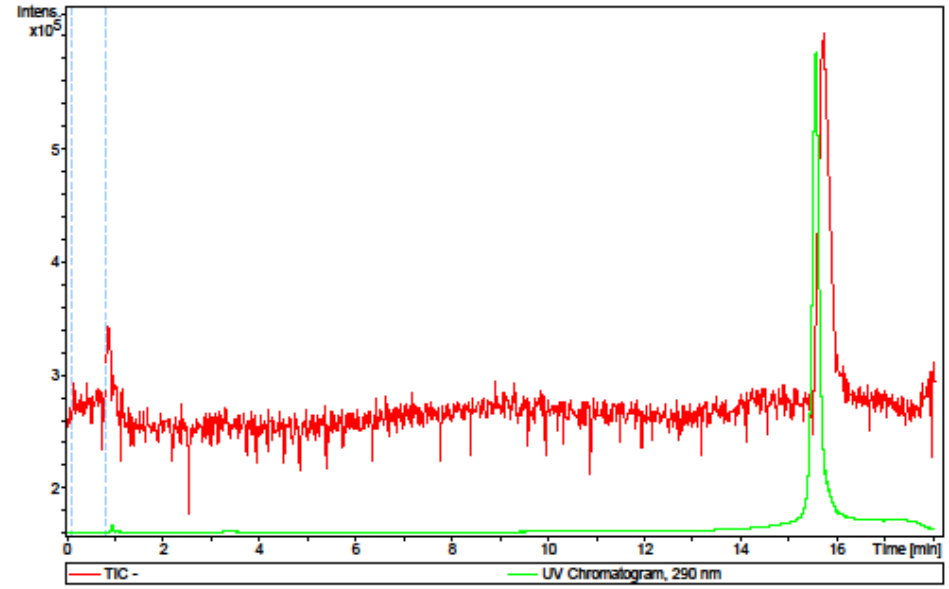
IC-5

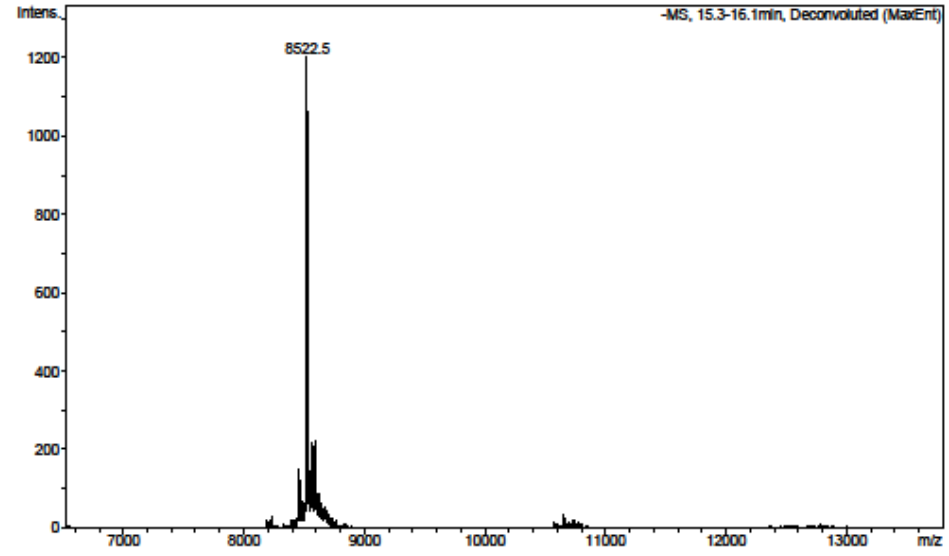
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp oligo generic hplc tune 08july15 retune.m	Acquisition Date	02/11/2015 15:30:16
Sample Name	C1114	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 02/11/2015 16:26:20 Page 1 of 1

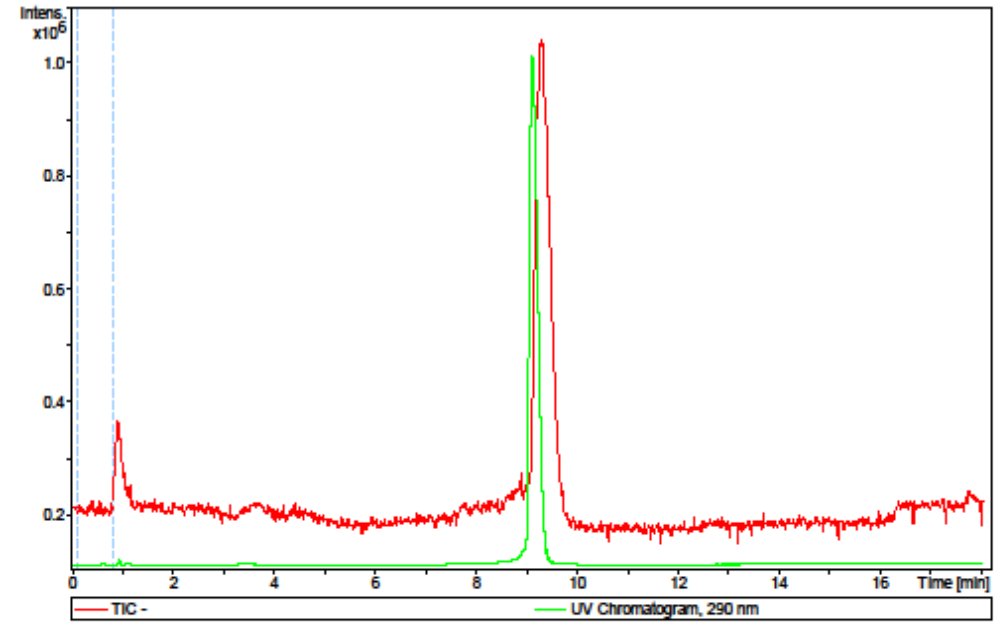
CT-mC

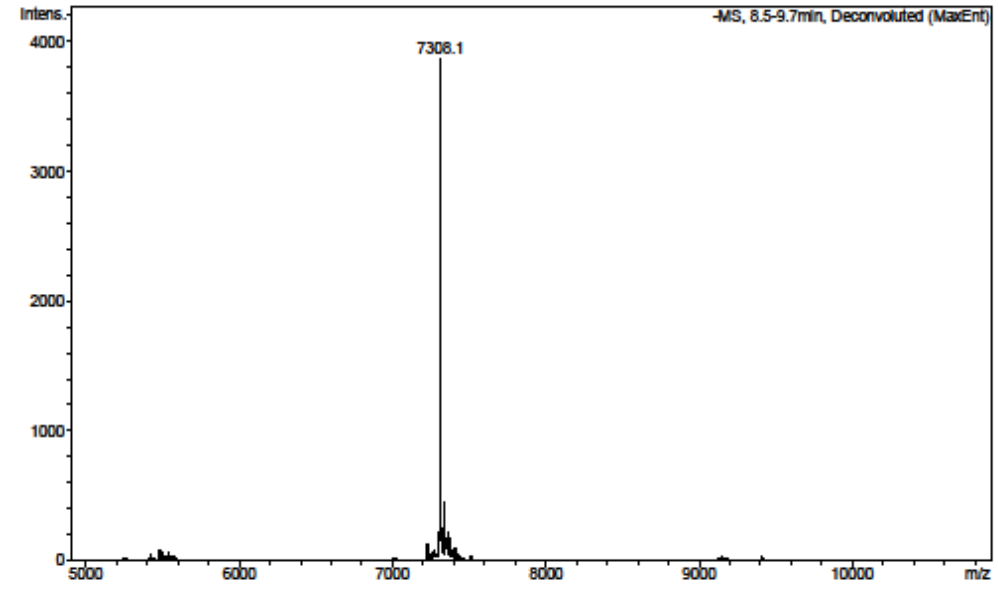
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp oligo generic hplc tune 08july15 retune.m	Acquisition Date	16/05/2016 13:30:28
Sample Name	C2321	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 17/05/2016 07:50:50 Page 1 of 1

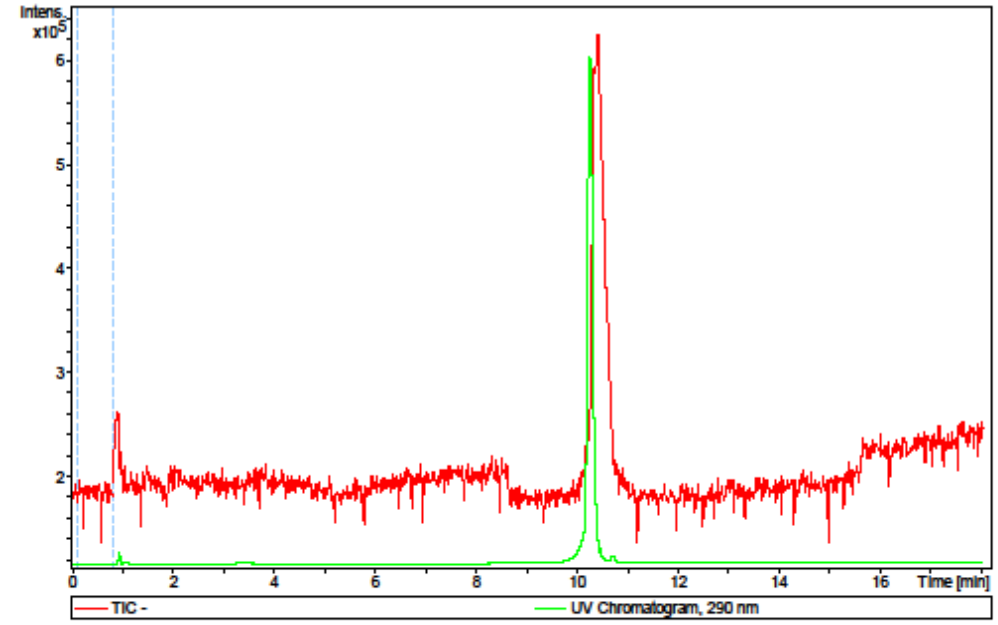
CT-1

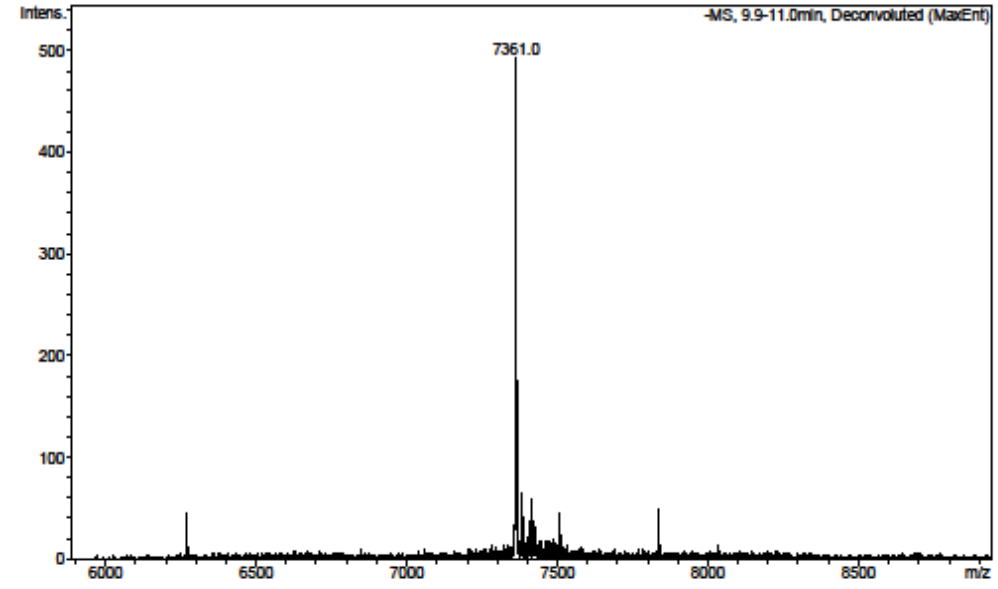
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp oligo generic hplc tune 08july15 retune.m	Acquisition Date	16/05/2016 13:49:27
Sample Name	C2322	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 17/05/2016 07:54:40 Page 1 of 1

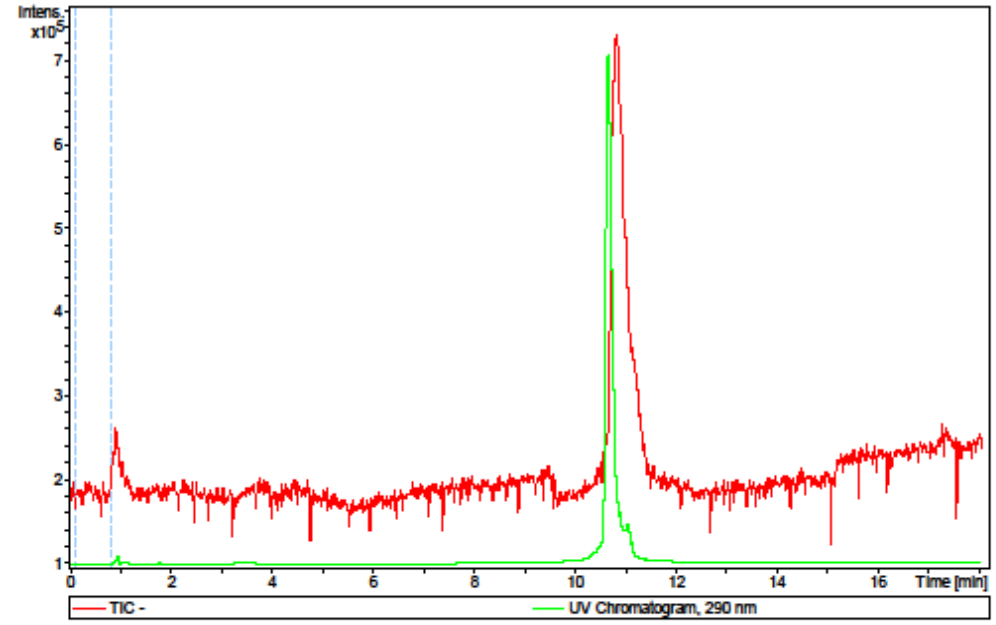
CT-2

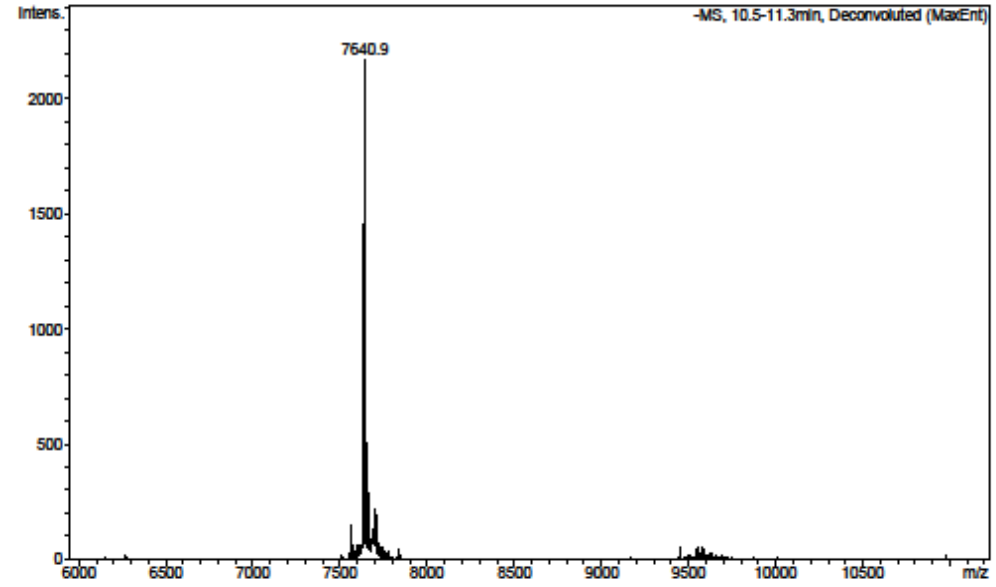
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp oligo generic hplc tune 08july15 retune.m	Acquisition Date	16/05/2016 14:08:28
Sample Name	C2323	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 17/05/2016 07:58:13 Page 1 of 1

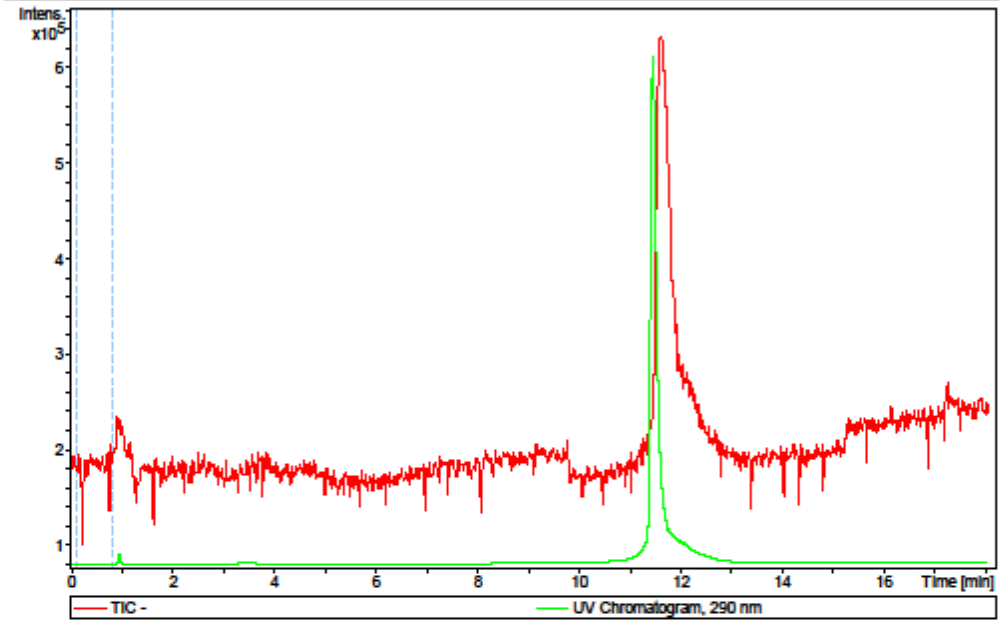
CT-3

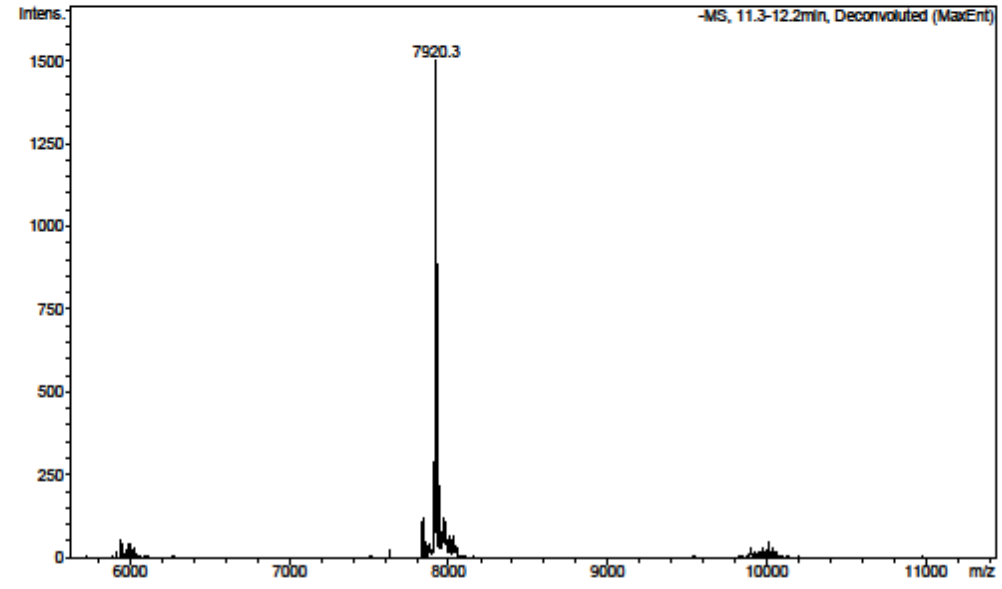
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp oligo generic hplc tune 08july15 retune.m	Acquisition Date	16/05/2016 14:27:34
Sample Name	C2324	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 17/05/2016 08:01:09 Page 1 of 1

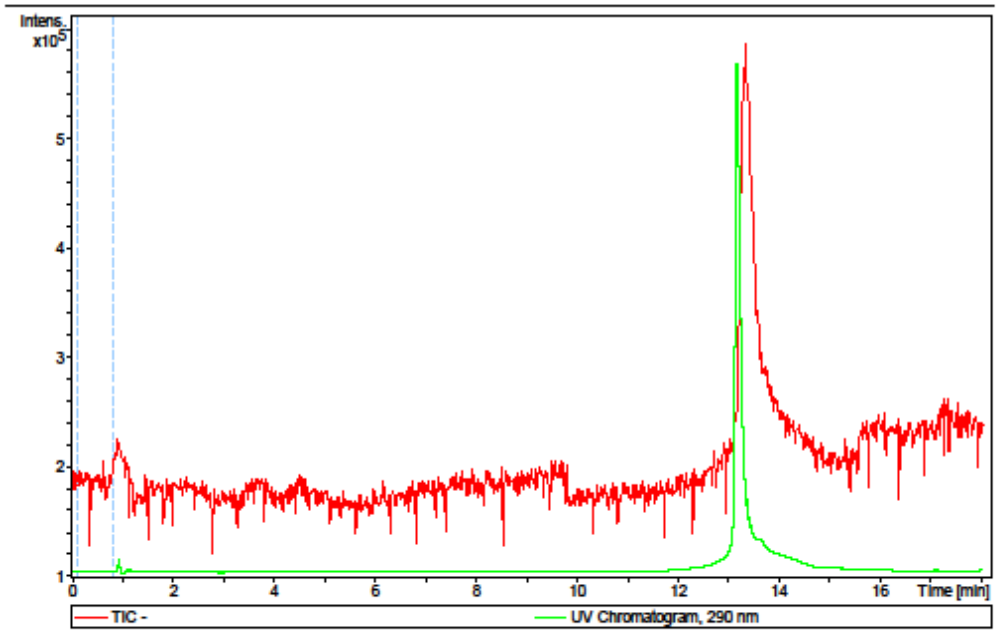
CT-4

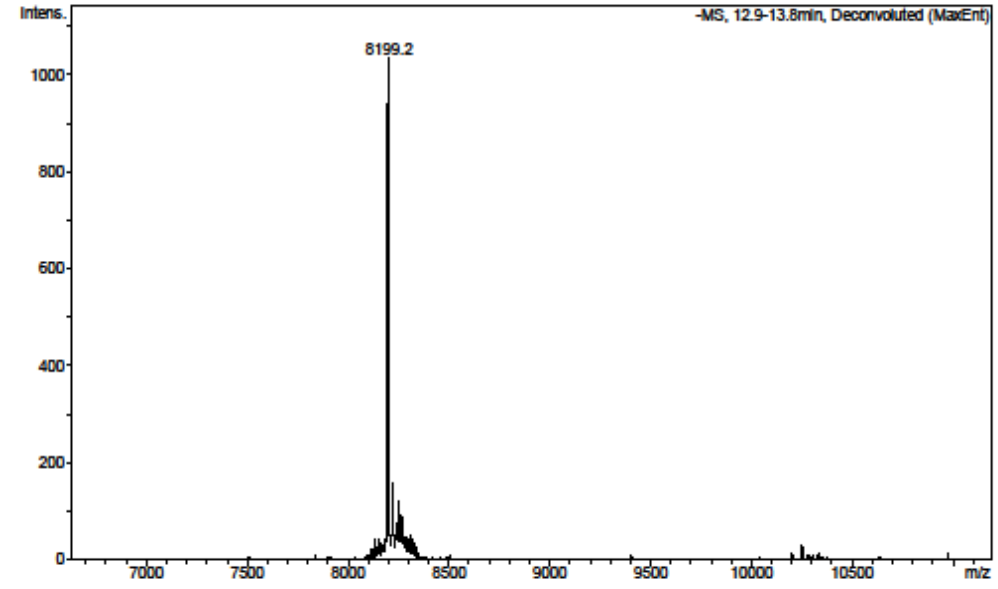
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp oligo generic hplc tune 08july15 retune.m	Acquisition Date	16/05/2016 14:46:32
Sample Name	C2325	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 17/05/2016 08:04:13 Page 1 of 1

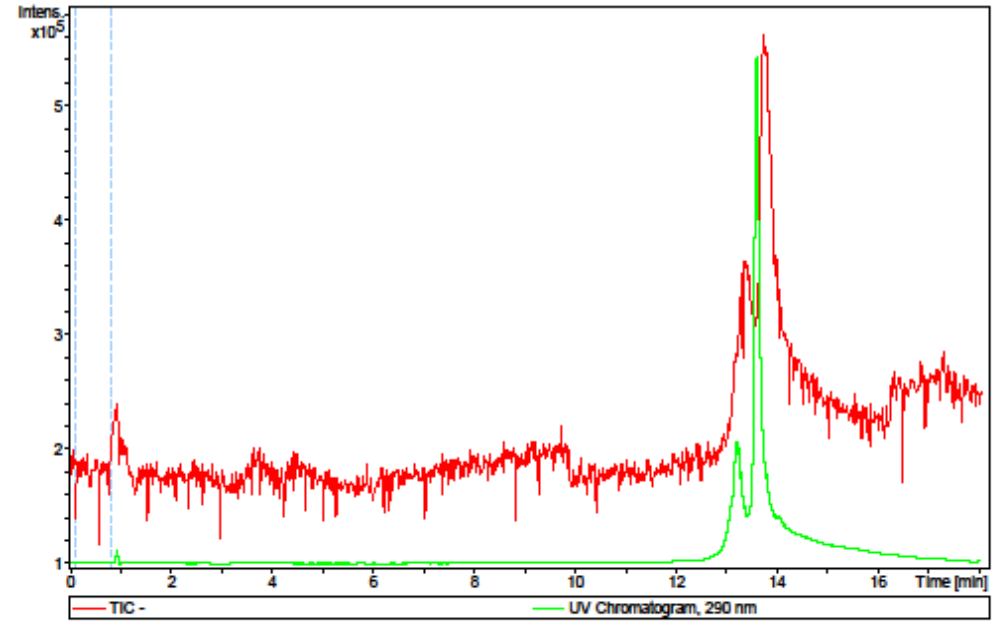
CT-5

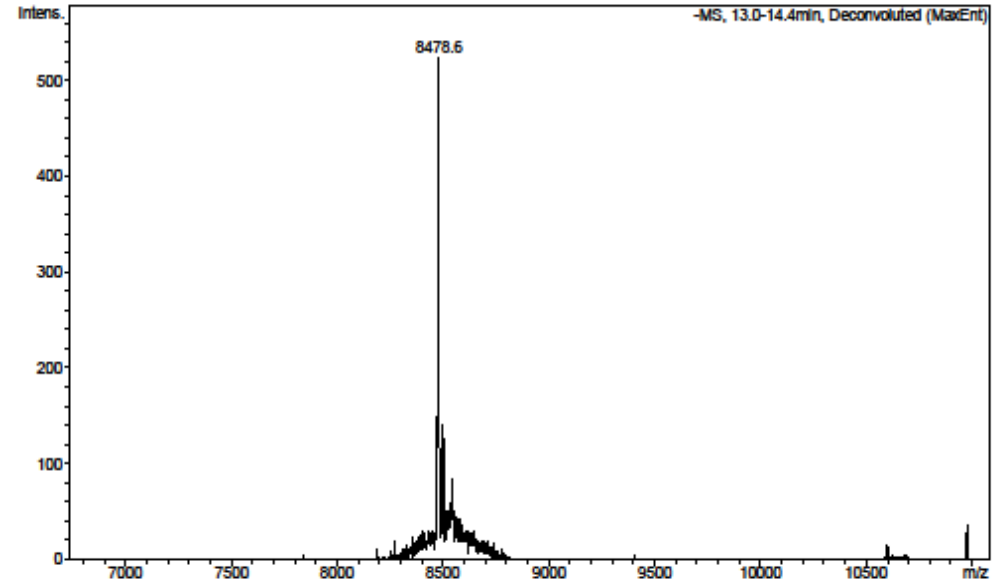
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp oligo generic hplc tune 08july15 retune.m	Acquisition Date	16/05/2016 15:05:34
Sample Name	C2326	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 17/05/2016 08:06:10 Page 1 of 1

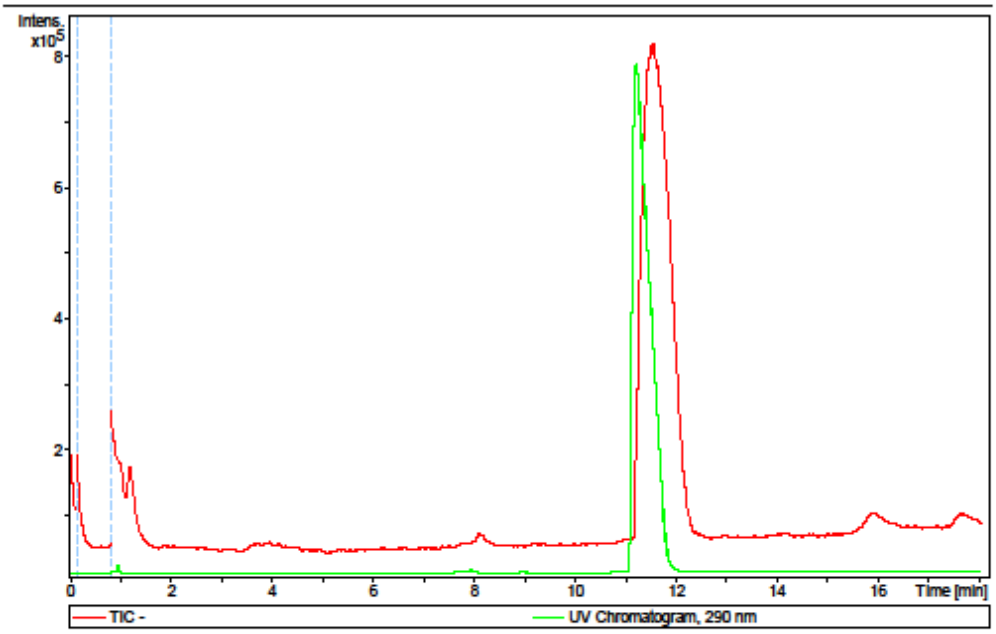
CT-dC

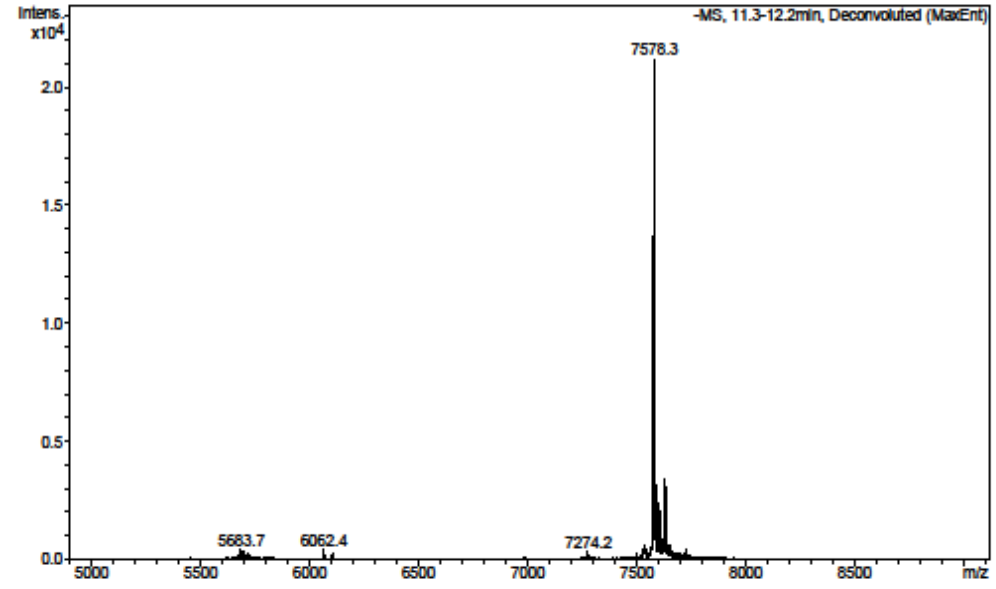
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp dg 6 oligo generic hplc tune 14mar17.m	Acquisition Date	18/07/2017 19:06:43
Sample Name	C5847	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 19/07/2017 10:04:53 Page 1 of 1

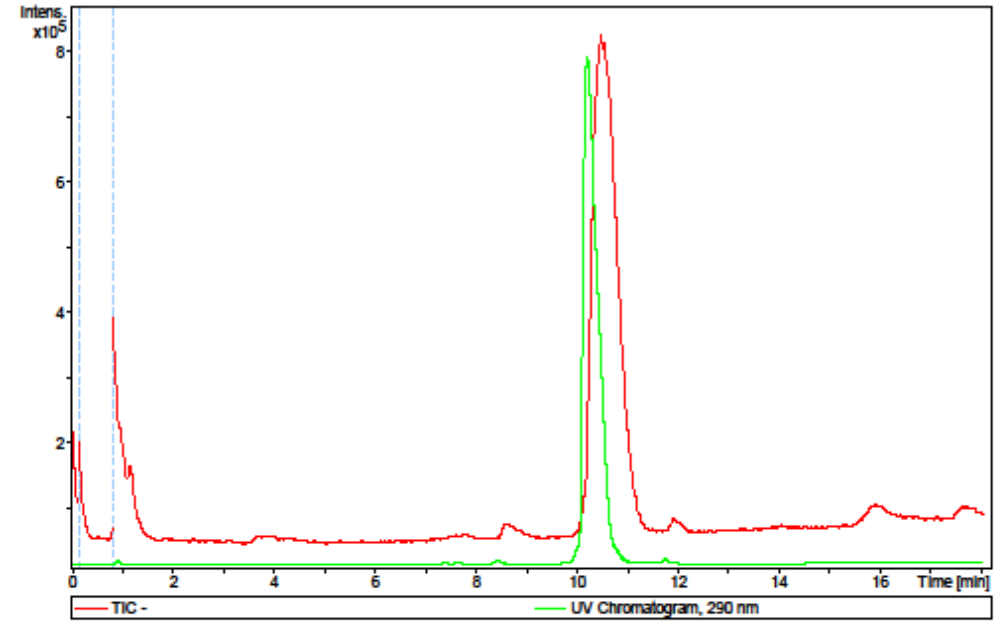
CT-6

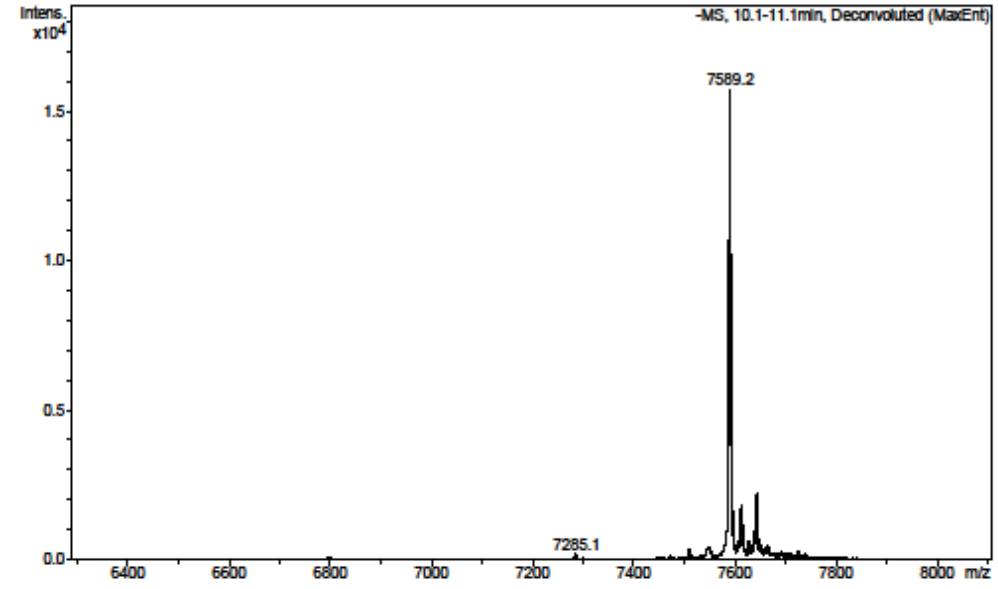
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp dg 6 oligo generic hplc tune 14mar17.m	Acquisition Date	18/07/2017 19:25:49
Sample Name	C5848	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 19/07/2017 10:08:57 Page 1 of 1

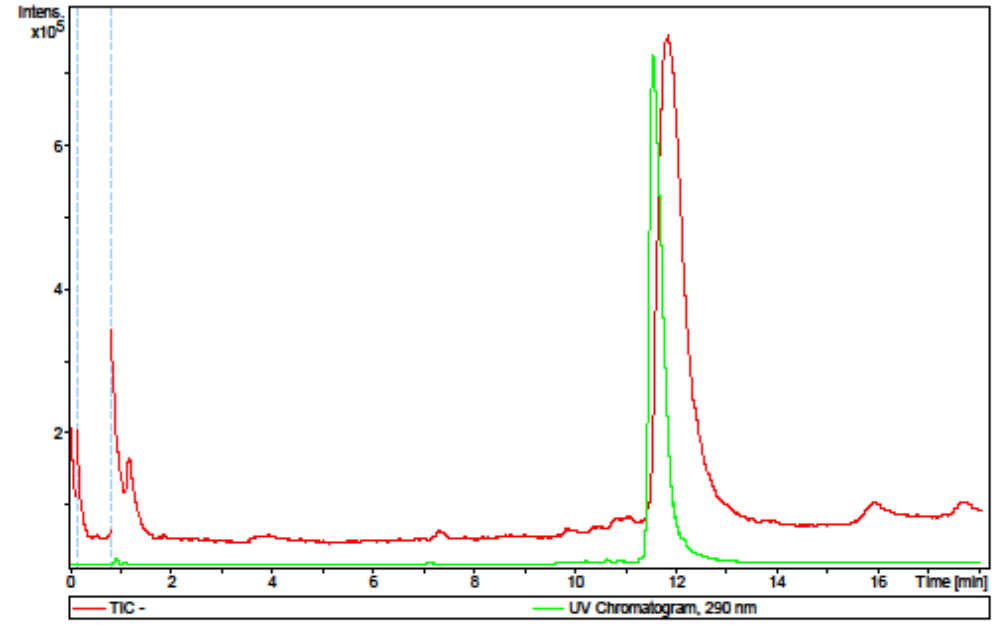
CT-7

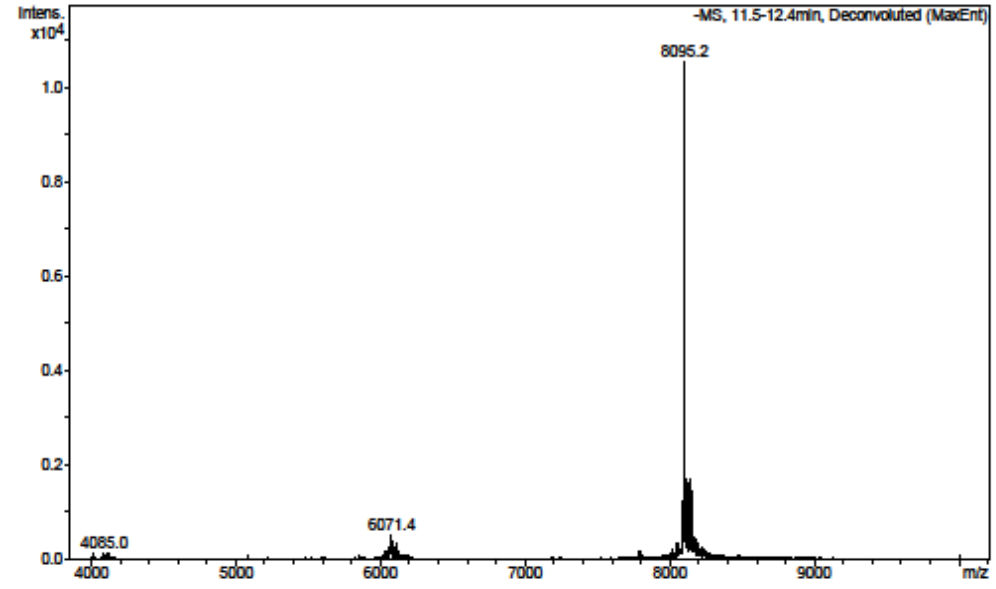
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp dg 6 oligo generic hplc tune 14mar17.m	Acquisition Date	18/07/2017 19:44:51
Sample Name	C5849	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 19/07/2017 10:14:28 Page 1 of 1

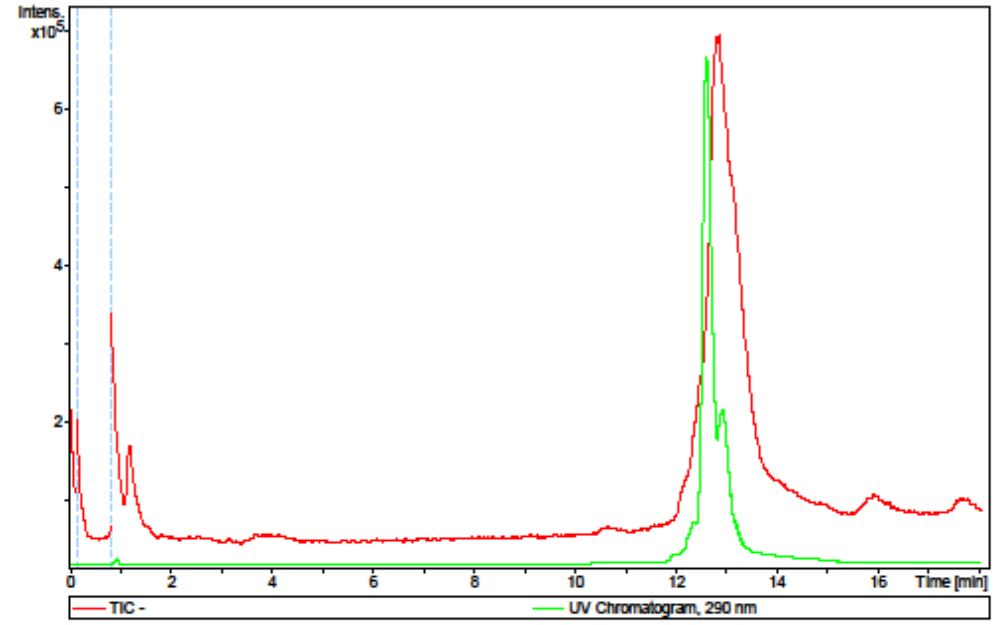
CT-8

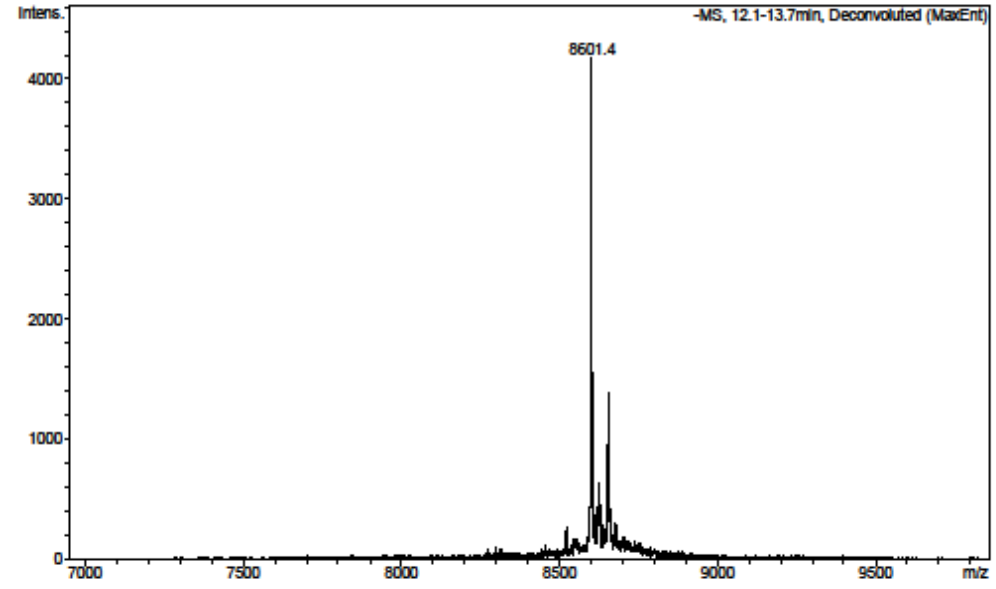
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp dg 6 oligo generic hplc tune 14mar17.m	Acquisition Date	18/07/2017 20:03:56
Sample Name	C5850	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 19/07/2017 10:20:37 Page 1 of 1

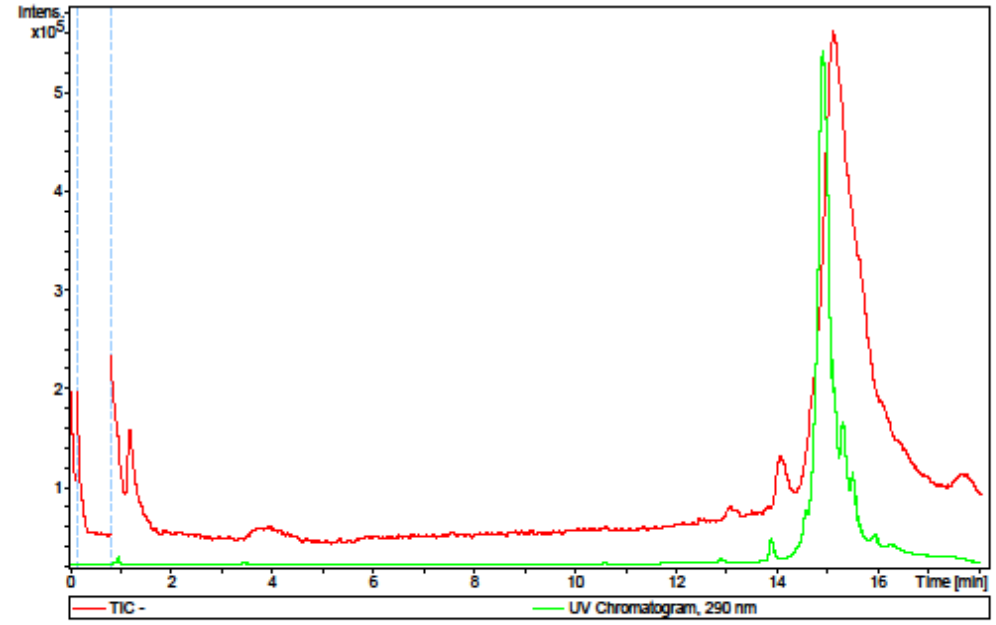
CT-9

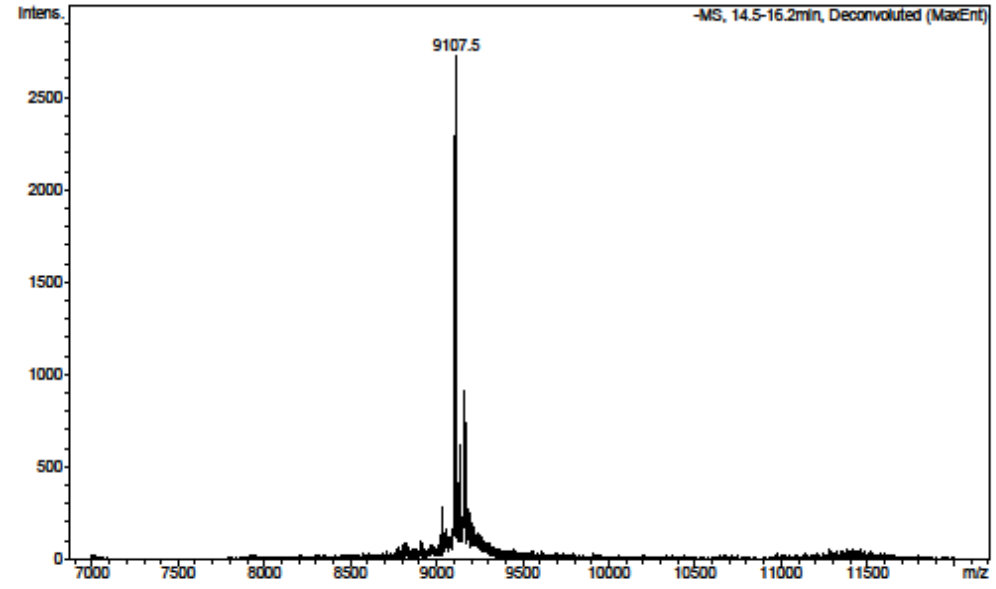
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp dg 6 oligo generic hplc tune 14mar17.m	Acquisition Date	18/07/2017 20:22:59
Sample Name	C5851	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 08/02/2018 09:18:01 Page 1 of 1

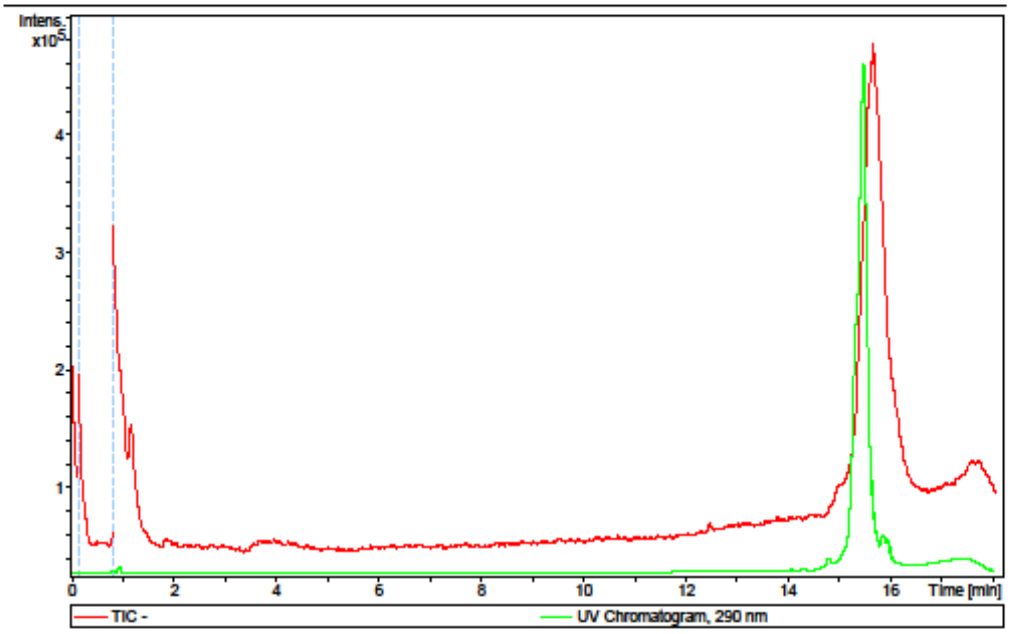
CT-10

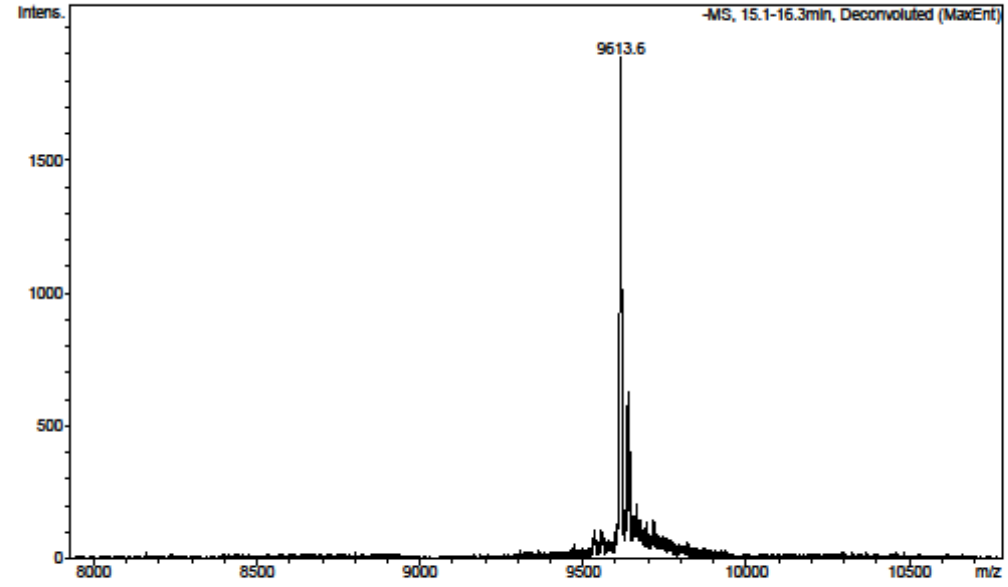
Mass Spectrometry Report

atdbio

Analysis Info

Method	300 high temp dg 6 oligo generic hplc tune 14mar17.m	Acquisition Date	18/07/2017 21:01:08
Sample Name	C5852	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 19/07/2017 10:31:45 Page 1 of 1

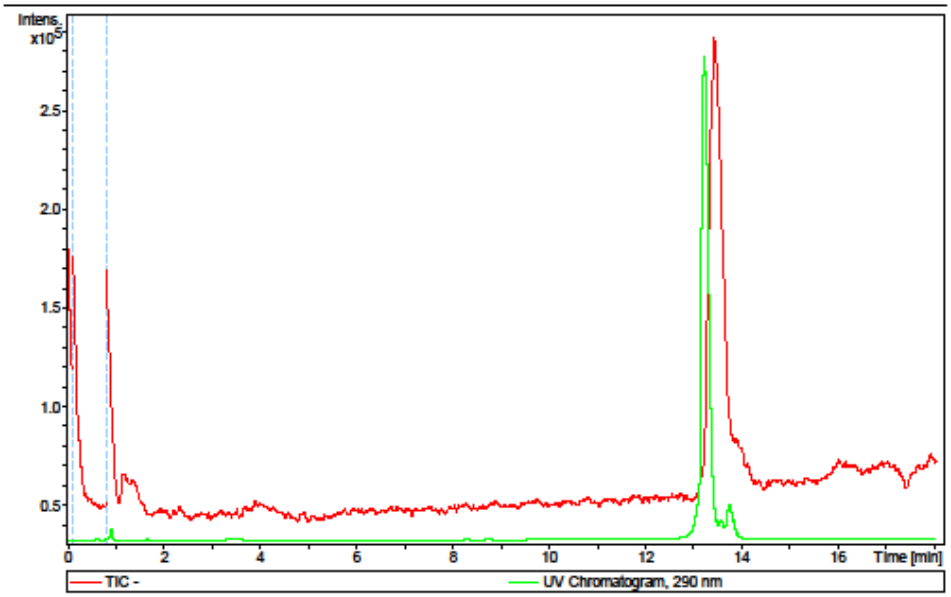
CT-11

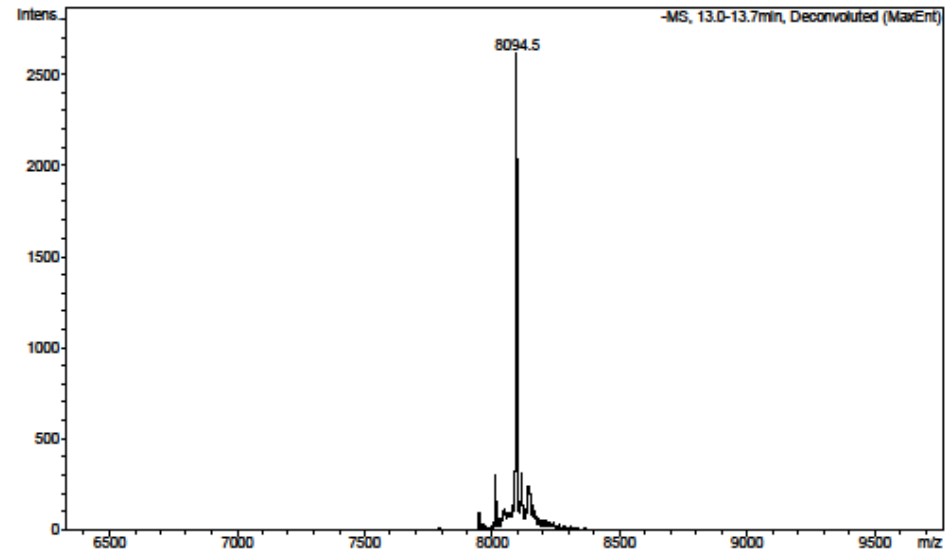
Mass Spectrometry Report

atdbio

Analysis Info

Method	320 high temp dg 8 oligo generic hplc tune 14mar17.m	Acquisition Date	28/11/2017 15:28:33
Sample Name	C6960	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 29/11/2017 08:09:13 Page 1 of 1

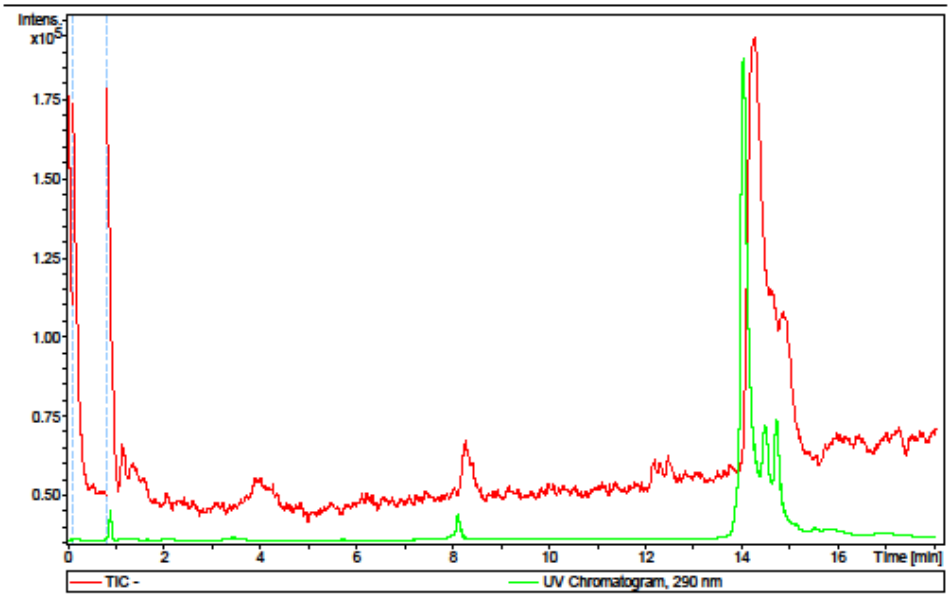
CT-12

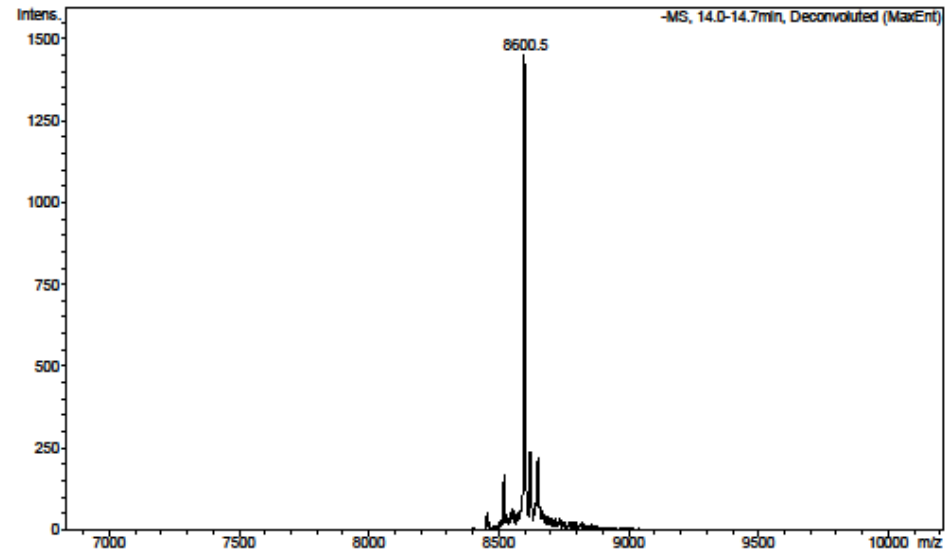
Mass Spectrometry Report

atdbio

Analysis Info

Method	320 high temp dg 8 oligo generic hplc tune 14mar17.m	Acquisition Date	28/11/2017 14:50:29
Sample Name	C6956	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 29/11/2017 08:04:29 Page 1 of 1

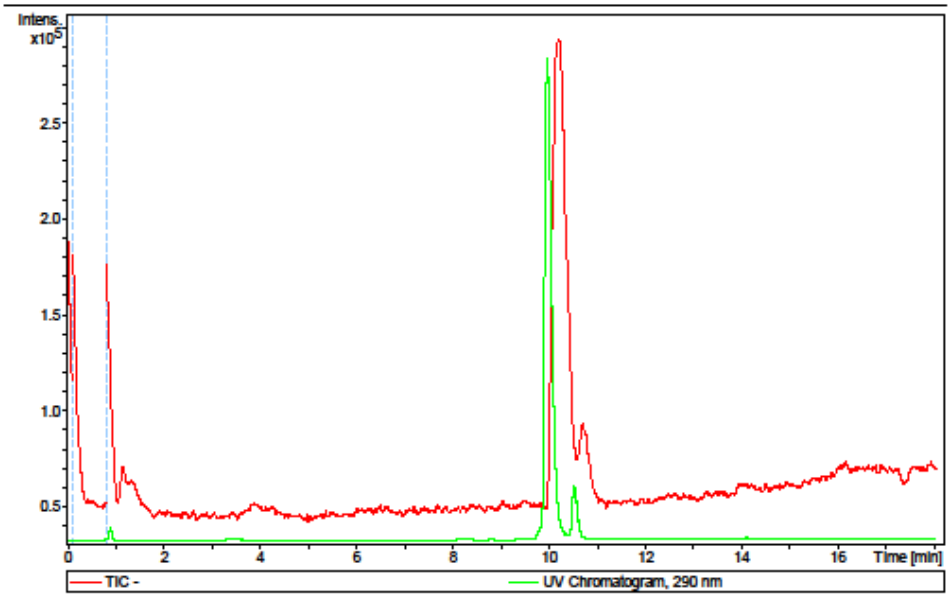
CT-13

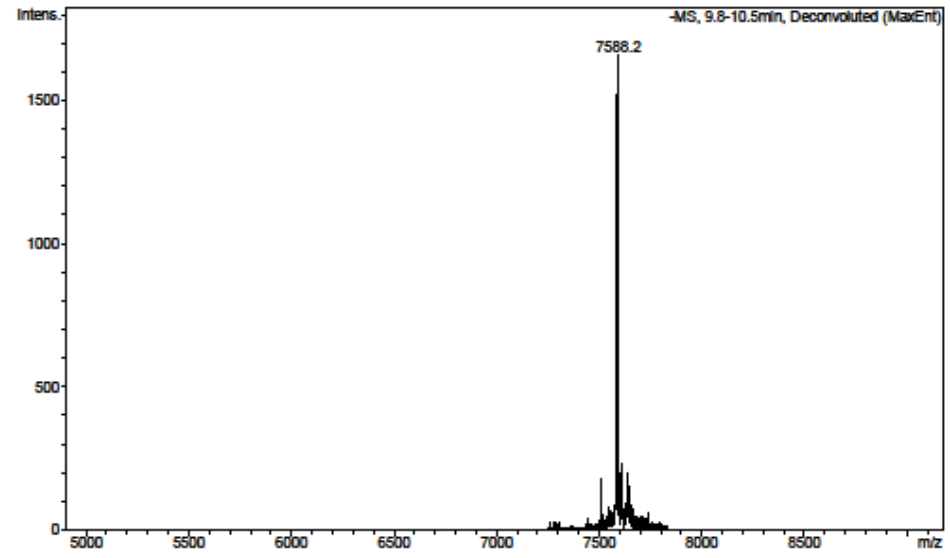
Mass Spectrometry Report

atdbio

Analysis Info

Method	320 high temp dg 8 oligo generic hplc tune 14mar17.m	Acquisition Date	28/11/2017 15:09:30
Sample Name	C6958	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 29/11/2017 08:06:52 Page 1 of 1

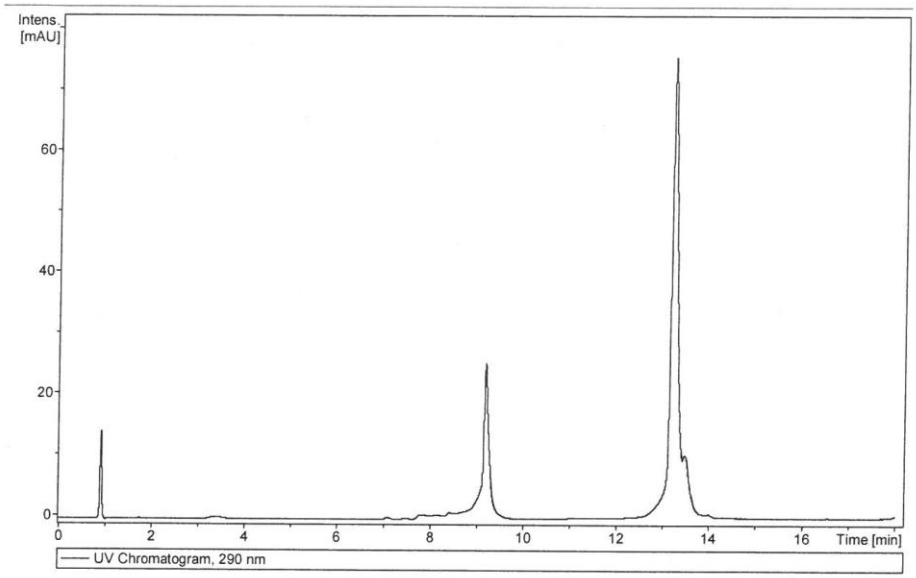
IC-098

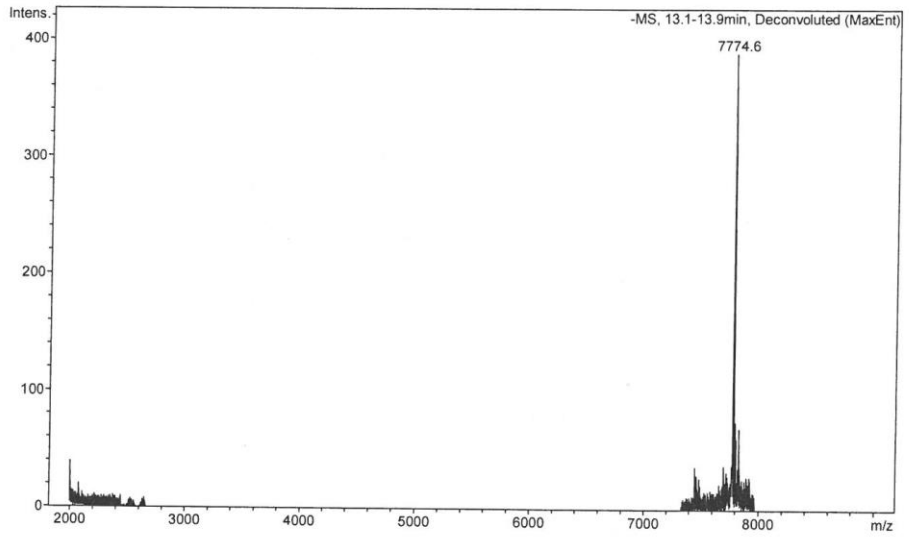
Mass Spectrometry Report

atdbio

Analysis Info

Method	320 high temp dg 8 oligo generic hplc tune 14mar17.m	Acquisition Date	09/02/2018 11:56:33
Sample Name	c7442 PK3	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 12/02/2018 10:56:21 Page 1 of 1

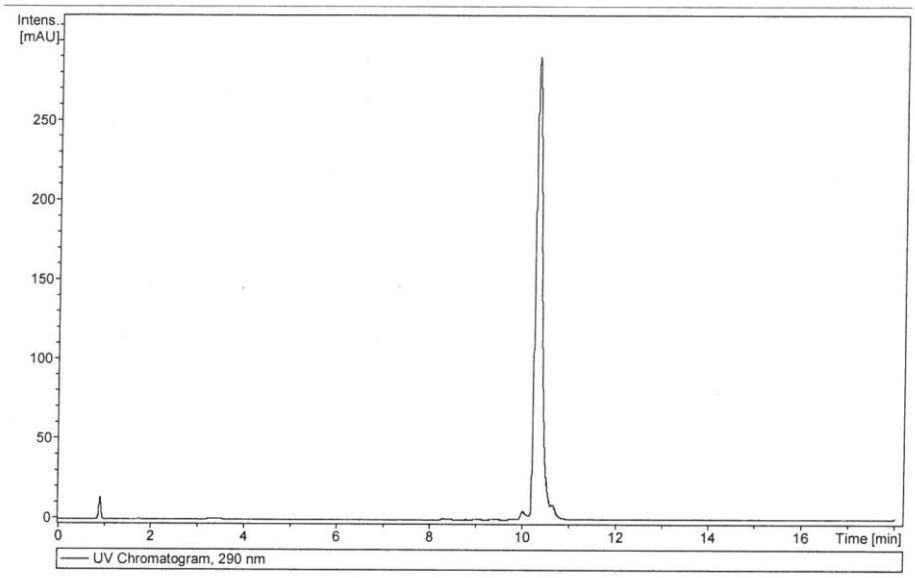
NG-437

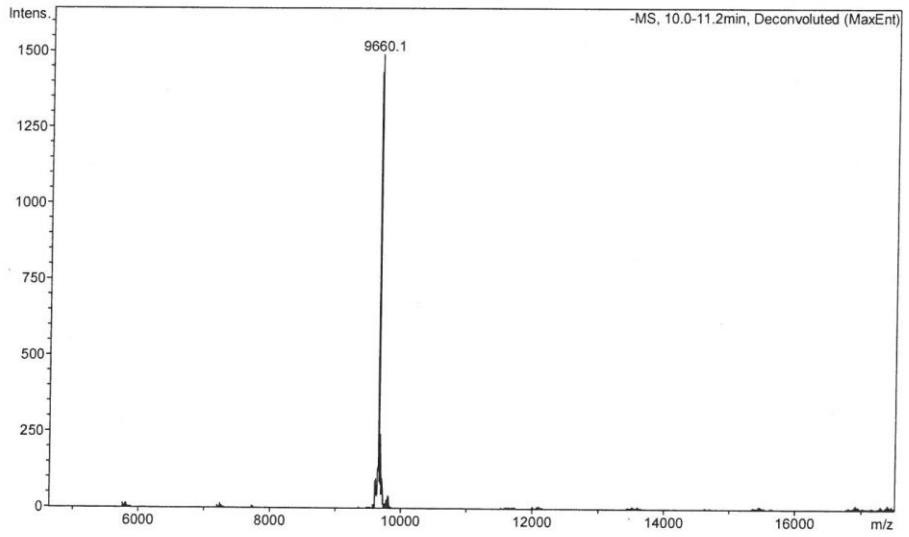
Mass Spectrometry Report

atdbio

Analysis Info

Method	320 high temp dg 8 oligo generic hplc tune 14mar17.m	Acquisition Date	09/02/2018 12:15:37
Sample Name	c7441	Operator	Bruker09
Comment		Instrument	micrOTOF





Bruker Compass DataAnalysis 4.0 printed: 12/02/2018 10:54:58 Page 1 of 1

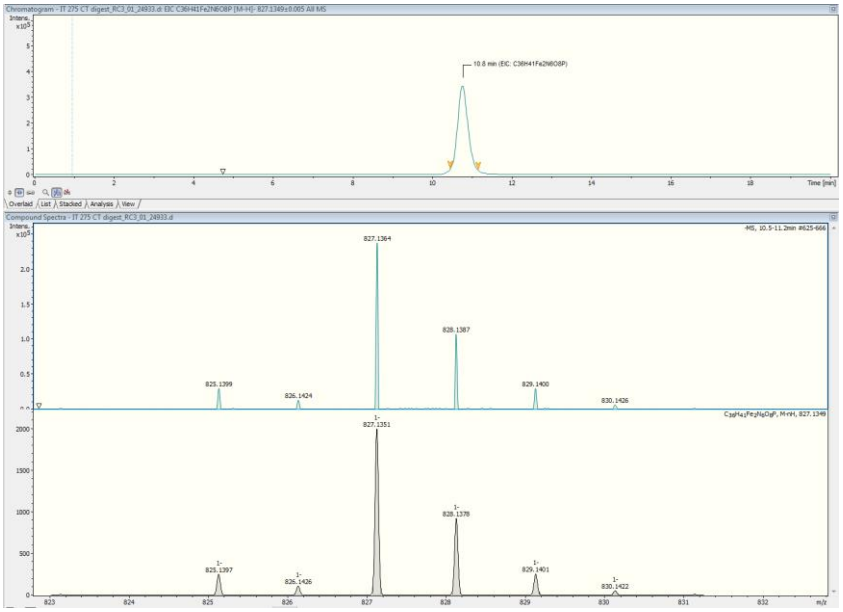
Appendix 6

HPLC-MS Traces of S1 nuclease digestions

CT-6

Digest to **113**

HRMS (ESI) calculated for $C_{36}H_{40}Fe_2N_6O_8P$ [M-H]$^-$: *m/z* 827.1350, found 827.1364.



CT-dC

Digest to **112**

HRMS (ESI) calculated for $C_{37}H_{46}Fe_2N_4O_7P$ [M-H]$^-$: *m/z* 801.1809, found 801.1813.